

## 04

한국문화예술위원회 예술정책연구

arts change the world

통합



문화이용권

빅데이터를 활용한

문화소외계층 문화

향유실태 분석 및

개선방안 연구 /

2018. 7

## 통합문화이용권 빅데이터를 활용한 문화소외계층 문화향유실태 분석 및 개선방안 연구

2018년 07월 인쇄

2017년 12월 발행

발행처 한국문화예술위원회

편집 한국문화예술위원회

한국문화예술위원회

58217 전라남도 나주시 빛가람로 640(빛가람동 352)

전화 061-900-2100, 2200

팩스 061-900-2362

홈페이지 [www.arko.or.kr](http://www.arko.or.kr)이메일 [arko@arko.or.kr](mailto:arko@arko.or.kr)

본 보고서는 무단전제를 금하며,

내용의 일부를 가공하거나 인용할 때에는

반드시 출전을 밝히시기 바랍니다.

# 통합문화이용권 빅데이터를 활용한 문화소외계층 문화향유실태 분석 및 개선방안 연구

/ 2018. 7



## 제 출 문

한국문화예술위원회 귀하

본 보고서를 통합문화이용권 빅데이터를 활용한 문화소외계층 문화향유실태 분석 및 개선방안 연구의 최종보고서로 제출합니다.

2018년 7월 18일

주 관 기 관 : 주식회사 선도소프트

연구책임자 : 주식회사 선도소프트 최 인 호

연 구 원 : 주식회사 선도소프트 윤 훈 주

이 강 근

박 주 혁

장 원 덕

박 미 라

최 연 지

김 지 원

조 영 호

조 은 경



# 목 차

## 제1장 서론

### 제1절 : 연구의 배경 및 필요성

1. 연구의 배경	18
-----------	----

### 제2절 : 연구의 목적

1. 연구의 목적	20
-----------	----

### 제3절 : 연구의 범위 및 내용

1. 연구의 범위	21
2. 연구의 내용	22

## 제2장 데이터 수집 및 정제

### 제1절 : 표준방법론

1. 표준방법론 적용	24
-------------	----

### 제2절 : 데이터 수집

1. 핵심 데이터의 이해	27
2. 데이터 수집	29

### 제3절 : 데이터 정제 및 구축

1. 데이터 정제 및 전처리	32
2. 데이터 구축	40

## 제3장 데이터 분석 및 알고리즘 개발

### 제1절 : 주제별 표준분석모델

1. 주제별 표준분석모델 정의	46
------------------	----

### 제2절 : 주제별 데이터 분석

1. 분석 시나리오 설정	53
2. 주제별 데이터 분석	55

제4장 결론 및 제언

제1절 : 결론

1. 연구의 결론	102
-----------	-----

제2절 : 제언

1. 정책 활용방안 및 제언	108
-----------------	-----

## 표 목 차

<표 1-1> 단계별 최종 및 세부 목표	21
<표 2-1> 요구사항 분석 및 적용방안 도출	25
<표 2-2> 문화누리카드 주요 내용	27
<표 2-3> 문화누리카드 신청 자격	28
<표 2-4> 문화누리카드 주요 사용처	29
<표 2-5> 데이터 수집 테이블 목록	30
<표 2-6> 통합문화사용권(문화누리카드) 데이터	31
<표 2-7> 식별자 예시	33
<표 2-8> 속성자 예시	33
<표 2-9> 비식별 조치 방법	34
<표 2-10> 비식별 조치 방법	35
<표 2-11> 데이터 전처리 시 고려사항	36
<표 2-12> 정제된 필드 목록	39
<표 3-1> '15년 기준 사용카드 수 가 많은 주민센터 (상위 20위)	64
<표 3-2> '16년 기준 사용카드 수 가 많은 주민센터 (상위 20위)	64
<표 3-3> '17년 기준 사용카드 수 가 많은 주민센터 (상위 20위)	65
<표 3-4> 발급 카드 수 대비 미사용 카드 수가 적은 지역 비교(높은 사용)	66
<표 3-5> '15년 기준 사용카드 수 가 적은 주민센터 (상위 20위)	68
<표 3-6> '16년도 기준 사용카드 수 가 적은 주민센터 (상위 20위)	68
<표 3-7> '17년도 기준 사용카드 수 가 적은 주민센터 (상위 20위)	69
<표 3-8> 발급 카드 수 대비 미사용 카드 수 비교 (낮은 사용)	70
<표 3-9> 문화누리카드 문의사항 테이블 데이터 구분 사항	74
<표 3-10> 단어 빈도 분석 소스	74
<표 3-11> 문화누리카드 문의사항 질문 유형	75
<표 3-12> 문화누리카드 문의사항 단어 빈도 분석 결과	76
<표 3-13> 사용분야(업종) 위치 단어 빈도 분석 결과	76
<표 3-14> 사용분야(업종) 사용 단어 빈도 분석 결과	77
<표 3-15> 기타 단어 빈도 분석 결과	78

<표 3-16> 발급자격 및 서류 문의사항 단어 빈도 분석 결과	79
<표 3-17> 비밀번호 변경 단어 빈도 분석 결과	80
<표 3-18> 시스템 오류, 잔액조회 단어 빈도 분석 결과	81
<표 3-19> 자부담 충전 단어 빈도 분석 결과	82
<표 3-20> 카드 분실 및 등록 단어 빈도 분석 결과	83
<표 3-21> 카드 합산 단어 빈도 분석 결과	84

## 그림 목 차

[그림 1-1] 빅데이터를 활용한 데이터 중심의 분석 기법	19
[그림 1-2] 빅데이터를 활용한 데이터 중심의 분석 방안	20
[그림 1-3] 빅데이터를 활용한 데이터 중심의 분석 방안	21
[그림 2-1] 표준방법론	24
[그림 2-2] 문화누리카드 지원대상	27
[그림 2-3] 문화누리카드 지원 대상	32
[그림 2-4] 데이터 삭제 및 마스킹	35
[그림 2-5] 프로세스 진행도	36
[그림 2-6] 데이터 구조화를 위한 입력데이터·처리 결과 데이터	37
[그림 2-7] 프로파일링 화면	37
[그림 2-8] 필드 오류 확인 절차	38
[그림 2-9] 비즈니스 규칙 확인	38
[그림 2-10] Null 규칙 검증 화면	39
[그림 2-11] 통합문화이용권(문화누리카드) 의사결정 모델을 위한 융합 데이터셋 구성(안)	40
[그림 2-12] 데이터 융합 예시 (1)	40
[그림 2-13] 데이터 융합 예시 (2)	41
[그림 2-14] 표준정규분포 : $Z = (x - \text{평균}) / \text{표준편차}$	42
[그림 2-15] 테스트 수행 결과	42
[그림 2-16] 통합된 데이터 정리화면	43
[그림 2-17] 구축에 활용한 카드 데이터 일부지역	44
[그림 2-18] 융합 데이터셋 구축(예_일부지역 대상 유동인구 DB)	44
[그림 3-1] 주제별 의사결정 표준분석모델 개념 정의	46
[그림 3-2] Linear Methods for Classification	47
[그림 3-3] Clustering Analysis	48
[그림 3-4] Anomaly detecion	48
[그림 3-5] Dimension reduction	49
[그림 3-6] Association Analysis	50
[그림 3-7] Sample Association Analysis	50

[그림 3-8] Decision Tree를 사용한 소비 예측	51
[그림 3-9] Random forests	51
[그림 3-10] Gradient boosting machine	52
[그림 3-11] Deep learning	52
[그림 3-12] 연령대별 카드 사용내역	55
[그림 3-13] 사용분야(업종)별 수혜자의 카드사용	56
[그림 3-14] 수혜자의 카드사용 금액과 사용 사용분야(업종) 내역	57
[그림 3-15] 분석용 데이터셋 구축 및 빅데이터 분석 방법론	57
[그림 3-16] 지역별 특성 도출	58
[그림 3-17] 계층그룹별 수혜자 분포 (2017)	59
[그림 3-18] 계층그룹별 수혜자 분포 (2014 ~ 2016)	59
[그림 3-19] 계층그룹 연령대별 수혜자 분포	60
[그림 3-20] 장애인 계층 사용분야(업종) 이용현황 분석	61
[그림 3-21] 사용 선호도 분석	62
[그림 3-22] 전체 연령대 분포	62
[그림 3-23] 발급 카드 대비 미사용 카드 수 비율	63
[그림 3-24] 지역 내 전체 사용분야(업종) 수와 지역구분별 평균 사용분야(업종) 비교분석 (카드 사용률이 높은 지역)	67
[그림 3-25] 지역 내 전체 사용분야(업종) 수와 지역구분별 평균 사용분야(업종) 비교분석 (카드 사용률이 낮은 지역)	71
[그림 3-26] 발급 카드 대비 사용건 수 비교	71
[그림 3-27] 상위 50% 재정자립도 상관관계	72
[그림 3-28] 상위 50% 재정자립도	72
[그림 3-29] 중위 50% 이하 재정자립도 상관관계	73
[그림 3-30] 중위 50% 이하 재정자립도	73
[그림 3-31] 전화문의사항(일부분)	79
[그림 3-32] (좌) 부산 전체 사용분야(업종) 분야별 비율(우) 주거인구와 사용분야(업종) 분포도	86
[그림 3-33] 동일 사용분야(업종)의 밀집도 분석	86
[그림 3-34] 부산 진구 일대 사용분야(업종) 분포	87
[그림 3-35] 문화일반 사용분야(업종) 분포 분석	88

[그림 3-36] (좌)숙박 사용분야(업종) 분석, (우) 도서 사용분야(업종) 분석	89
[그림 3-37] 이상결제내역 모니터링 분석 방향 설정	90
[그림 3-38] Anomaly Detection 알고리즘	90
[그림 3-39] 이상결제내역 모니터링 방법론 (1)	91
[그림 3-40] 이상결제내역 모니터링 방법론 (2)	91
[그림 3-41] 보조금 금액 상향효과 분석 방법론	93
[그림 3-42] 평균 사용 건수 / 사용카드 수(한 카드당 평균 사용건수)	94
[그림 3-43] 평균 사용 금액 / 사용카드 수(한 카드당 평균 사용 금액)	94
[그림 3-44] 평균 1건 당 카드사용 금액	94
[그림 3-45] 사용분야(업종)에 따른 카드사용내역 특성분석(지역별, 연도별)	95
[그림 3-46] 사용분야(업종)에 따른 카드사용내역 특성분석(연도별 증가율)	96
[그림 3-47] 사용분야(업종)별 평균 카드사용 금액	97
[그림 3-48] 장애인 계층 연도별 소비 사용분야(업종) 분포(2014~2017)	98
[그림 3-49] 기초수급자 계층의 연도별 소비 사용분야(업종) 분포(2014~2017)	99
[그림 3-50] 차상위 계층의 연도별 소비 사용분야(업종) 분포(2014~2017)	100
[그림 4-1] 융합 데이터셋 구축	103
[그림 4-2] 주제별 의사결정 표준분석모델 개념 정의	104
[그림 4-3] 수혜자 빅데이터 분석 방법론	104
[그림 4-4] 수혜자 분석 결론	105
[그림 4-5] 사용분야(업종) 분석 결론	105
[그림 4-6] 이상결제 내역 모니터링 분석 결론	106
[그림 4-7] 보조금 금액 상향 효과 분석 결론	107



# 제1장

---

## 서론

제1절 : 연구의 배경 및 필요성

1. 연구의 배경

제2절 : 연구의 목적

1. 연구의 목적

제3절 : 연구의 범위 및 내용

1. 연구의 범위

2. 연구의 내용

## 제1절 연구의 배경 및 필요성

## 1. 연구의 배경

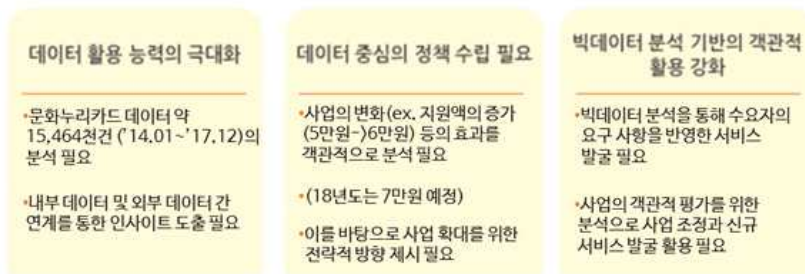
## □ 객관적이고 과학적인 “수혜자 맞춤형 서비스 제공” 정책수립 토대 마련

- 문화기본법 제 4조에 “모든 국민은 성별, 종교, 인종, 세대, 지역, 사회적 신분, 경제적 지위나 신체적 조건 등에 관계없이 문화 표현과 활동에서 차별을 받지 아니하고 자유롭게 문화를 창조하고 문화 활동에 참여하여 문화를 향유할 권리(이하 “문화권”이라 한다)를 가진다.”라고 명시되어 있음
- 이는 문화를 향유 하는 것은 국민의 권리로 보고 있다는 것이며, 또한, 헌법 제 11조 “누구든지 성별·종교 또는 사회적 신분에 의하여 정치적·경제적·사회적·문화적 생활의 모든 영역에 있어서 차별을 받지 아니한다.”라고 명시되어 있어 문화 향유에 중요성을 강조하고 있음
- 이러한 관점으로 문화는 모든 국민이 누려야 할 권리이고 문화생활 영역에 있어서 어느 누구도 소외 받으면 안 된다는 것임
- 즉, 개인의 문화적 향유는 생존권과 사회보장권 등과 같이 하나의 권리로 인식되고 있으며, 더 이상 문화생활을 누리는 것이 경제적으로 여유로운 사람들만의 것이 아니라 모든 국민이 당연히 보장받아야 된다는 것으로 이해할 수 있음(문보영, 2017)
- 통합문화이용권(문화누리카드)은 이런 시대 흐름에 따라 사회적·경제적 어려움으로 인해 문화예술을 향유하지 못하고 소외계층에게 다양한 문화예술여행·체육활동을 즐길 수 있는 문화누리카드를 지원하고, 이를 통해 소외계층의 문화향유기회 격차 해소 및 삶의 질 개선에 기여하고자 제공하는 바우처 사업으로 전국 161만 명의 국민을 대상으로 하는 국내 최대의 문화 복지 사업임
- 본 통합문화이용권(문화누리카드)<sup>1)</sup> 연구는 공공문화정보 중에서 거의 유일하게 개인정보와 문화소비에 대한 구체적인 데이터를 체계적으로 구축하고 있는 분야이나, 총괄적인 데이터 분석 이외에 빅데이터 분석기법을 이용한 분석과 활용에 대한 계획은 아직 검토되지 못하고 있는 실정임

1) **통합문화이용권(문화누리카드)**: 경제적 여건 등으로 인해 일상에서 문화를 누리기 어려운 분들에게 공연·전시·영화·도서 등 다양한 문화예술 프로그램, 국내 여행 및 체육 활동비용을 지원함으로써 소외계층에게 문화 향유 기회를 제공하는 문화 복지 사업

- 빅데이터 분석을 통해 수혜자의 요구사항을 반영한 서비스를 발굴할 필요가 있으며, 사업의 객관적 평가를 위한 분석을 통해 신규 서비스 발굴 활용의 필요성이 증가하면서 빅데이터 분석 기반의 객관적 활용 방안이 강조되고 있음
- 빅데이터 분석 기법을 활용해 수혜자의 요구사항에 대한 객관적이고 과학적인 ‘수혜자 맞춤형 서비스 제공’ 정책 수립의 토대 마련이 필요하며, 통합문화이용권 (문화누리카드) 사용 형태, 사업의 효과성 연구, 편의 증진, 부가 콘텐츠 개발 등 기초 자료로서의 객관적 자료가 필요함
- 또한 사업의 변화(5만원 → 6만원 지원액의 증가)등의 효과를 객관적으로 분석할 필요성과 이를 바탕으로 전략적 사업 확대를 위한 구체적 방향을 제시할 수 있는 데이터 중심의 정책 수립 필요성이 증가하고 있음
- 단순통계나 사례중심의 서비스로 수집되고 있는 데이터의 활용이 미비하기 때문에 서비스 기술적 한계를 극복하기 위한 방안이 필요함

[그림 1-1] 빅데이터를 활용한 데이터 중심의 분석 기법



## 제2절 연구의 목적

### 1. 연구의 목적

□ 본 연구는 빅데이터를 활용한 문화소외계층 문화향유 실태 분석 및 개선방안 도출을 주목적으로 하며, 구체적인 연구목적은 다음과 같음

- **지역별 특성을 반영한 사용분야(업종) 확대 및 실수혜자 사용을 증대 방안 제언**
  - (수혜자) 카드 승인 중심의 데이터를 수혜자 중심의 카드 데이터로 전환
  - (사용분야(업종)) 전국 사업체 총 조사 데이터를 활용한 시군구별 사용분야(업종) 비교
- **부정수급방지를 위한 예방적 모니터링 알고리즘 적용 방안 제언**
  - Anomaly Detection (이상치 분석) 알고리즘 모델링
- **예산 증액 등 수혜자 맞춤형 서비스 제공을 위한 의사결정 및 정책 수립 제언**
  - 보조금 금액 상향('16년 5만원 → '17년 6만원) 효과 분석

[그림 1-2] 빅데이터를 활용한 데이터 중심의 분석 방안

분석 분야	분석 내용
<b>1. 수혜자 분석</b> → - 수혜자 분석(자격, 나이 등) - 카드발급, 사용건수, 미사용 내역	<b>수혜자</b> <b>카드 승인 중심의 데이터를 수혜자 중심의 카드 데이터로 전환</b> · 기존 시도별 특징 분석에서 벗어나 지역별 특성을 고려하여 군점분석(시, 군, 광역시, 특별시 단위)을 통한 합리적 분석방법 제시
<b>2. 가맹점 분석</b> → - 사용분야(업종) 위치, 분포 - 사용분야(업종) 입지 검토	<b>가맹점</b> <b>전국 사업체 총 조사 데이터를 활용한 시군구별 가맹점 비교</b> · 외부데이터와 융합(통계청 거주인구데이터, 유동인구, 카드매출 등)을 통해 가맹점의 입지 분석 · 문화누리카드 사용 수 대비 문화누리 가맹점 비율을 통한 가맹점 상대적 분포 비교
<b>3. 이상결제내역 모니터링</b> → - 사용건수, 지역 - 과거 거래 패턴	<b>이상결제</b> <b>Anomaly Detection(이상치 분석)알고리즘 패턴</b> · 하나의 가맹점에서 분 단위 여러 건 카드결제가 일어난 패턴
<b>4. 보조금 상향 효과 분석</b> → - 카드사용내역 - 개인 충전액 사용 패턴	<b>보조금</b> <b>보조금 금액 상향('16년 5만원 → '17년 6만원) 효과 분석</b> · 보조금 상향 변화에 따른 가맹점, 지역, 발급자 특성 등 이용 행태 분석 · 가맹점 구분에 따른 개인 충전금 사용액, 나이별, 발급 자격에 따른 개인 충전액 사용 패턴 분석

□ 위와 같은 목적으로 한국문화예술위원회가 운영하는 통합문화이용권(문화누리카드) 사업과 관련한 데이터의 가치를 높이고, 나아가 능동적 정책수행기관으로서 역할을 수행하기 위한 한국문화예술위원회의 새로운 성장 동력의 발굴을 목적으로 함

- 즉, 과거 한국문화예술위원회의 통합문화사용권(문화누리카드) 자료에 대한 단순 통계정보를 빅데이터 정보 기반으로 변환시킴으로써 수혜자들의 사용분야(업종) 확충과 수혜자 맞춤형서비스 제공을 객관적이고, 체계적으로 할 수 있게 하고자 함

### 제3절 연구의 범위 및 내용

## 1. 연구의 범위

### □ 연구의 세부적 범위

- 문화소외계층 문화향유실태 분석 및 개선방안 도출을 위한 통합문화사용권(문화누리카드)의 빅데이터를 활용한 데이터의 효율적인 구축, 분석 모델 개발과 활용을 위해 전국을 대상으로 분석함

[그림 1-3] 빅데이터를 활용한 데이터 중심의 분석 범위

시간적 범위	▶ '14. 1.1 ~ '17.12.31 (단 분석의 효율성을 위해 외부 데이터와 시간 범위조정 가능)
내용적 범위	▶ 통합문화이용권 시스템 내 보유 자료를 바탕으로 외부 데이터 연계 분석
공간적 범위	▶ 전국

### □ 각 단계 최종 목표별 연구 추진 범위

<표 1-8> 단계별 최종 및 세부 목표

목표 구분	최종목표	세부목표
수집	통합문화사용권(문화누리카드) 시스템 내 데이터 및 외부 데이터 연계 데이터셋 구축	카드 결제 사용건 결제내역(수혜자별, 사용분야(업종)별)의 수집 및 정제, 외부 카드 데이터와 연계한 Mash up 데이터셋 구축
분석	특성 및 요인분석	융합 데이터를 바탕으로 수혜자의 활용 형태 분석, 사용분야(업종)별·지역별 가맹점의 사용 등 이상치 패턴 분석
예측	빅데이터 기반 수혜자 활용 알고리즘 개발	수혜자의 지역별·연령별·계층별 수요와 공급의 변화 예측
검증	시스템 로드맵 제시와 검증	지속적 데이터 수집방안 및 알고리즘 활용 방안 제시, 데이터 분석 결과의 검증, 활용확대를 위한 ICT 활용방안 제시
활성화	데이터 기반 의사결정 지원	통합문화사용권(문화누리카드) 사용률 확대를 위한 맞춤형 가맹점 및 수혜자 최적화 방안 제시, 지원액 증가에 따른 객관적 변화 예측

## 2. 연구의 내용

### □ 통합문화이용권(문화누리카드) 수혜자의 문화향유 실태조사 및 심층 분석

- 수혜자의 사용분야(업종)와 사용형태 등 분석 및 빅데이터 활용 접근 방법 제시
- 통합문화이용권(문화누리카드) 가맹점의 사용처·계절·연령·계층·사용요인·분야(도서, 공연, 영화 등) 사용금액 등 시스템 제공 자료 및 현장실사 조사 결과를 활용한 다각도 관점에서의 형태 분석
- 지원액 증가(5만원→6만원)에 따른 사용형태 변화 분석 및 지속적 증가 예상('17년 6만원→'20년 10만원/1인)에 따른 효과분석
- 통합문화이용권(문화누리카드) 발급 후 미사용자에 대한 지역별·연령별·계층별 상세 원인 분석 및 사용촉진을 위한 서비스 개발 방안 제시
- 최신 IT기기를 활용한 홍보 및 사용률 증대 방안 제시
- 대내외 시스템 연계를 통한 서비스 확대 방안 제시

### □ 연구 분석 결과 도출

- 지역별 특성을 반영한 가맹점 확대 및 실 수혜자 사용률 증대 방안 제언
- 부정수급 방지를 위한 예방적 모니터링 알고리즘 적용 방안 제언
- 예산 증액 등 수혜자 맞춤형 서비스 제공을 위한 의사결정 및 정책 수립 제언
- 향후 통합문화이용권(문화누리카드)사업 활성화 방안 및 효과성 제고를 위한 빅데이터 활용방안 및 개선과제 제언

## 제2장

---

### 데이터 수집 및 정제

제1절 : 표준방법론

1. 표준방법론 적용

제2절 : 데이터 수집

1. 핵심 데이터의 이해
2. 데이터 수집

제3절 : 데이터 정제 및 구축

1. 데이터 정제 및 전처리
2. 데이터 구축

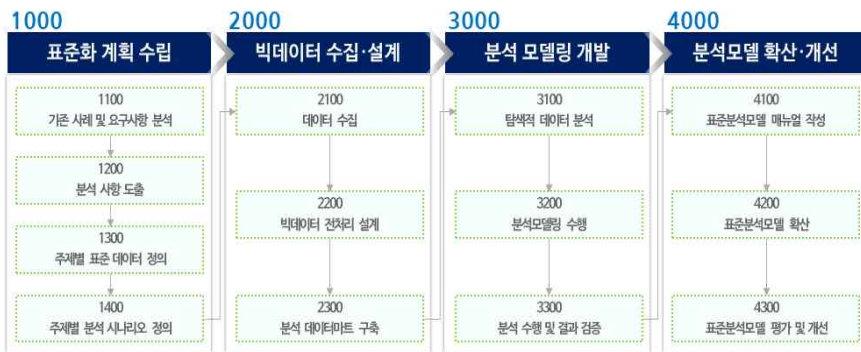
## 제1절 표준방법론

## 1. 표준방법론 적용

## □ 분석 모델링 표준 방법론 단계에 따른 작업 구성

- 분석 모델링 표준 방법론은 표준화 계획 수립, 빅데이터 수집·설계, 분석 모델링 개발, 분석모델 확산·개선의 단계로 총 4단계로 이루어짐

[그림 2-1] 표준방법론



- 표준화 계획 수립 단계는 기존 사례 분석에 의한 표준화 요구사항 정의 및 현황분석을 통하여 명확한 분석 주제 도출 및 분석에 필요한 데이터 항목과 분석 시나리오를 정의하는 단계임
  - 기존 사례 및 요구사항에 대한 분석을 시작으로 분석 사항을 도출하고 도출된 분석 사항에 대해 주제별 표준 데이터를 정의함
  - 주제별 분석 시나리오를 정의함으로써 다음 빅데이터 수집 및 설계 단계로 진행됨
- 빅데이터 수집 및 설계 단계에서는 빅데이터 분석을 위하여 필요한 데이터를 수집하고, 가공·변환하여 데이터의 저장을 위한 설계와 분석 데이터 마트를 생성함
- 수집된 데이터를 통해 분석 모델링 개발 단계에서 탐색적 데이터 분석을 통해 데이터의 특성을 파악하고, 분석 프로세스 기반 분석 모델링 및 수행을 통해 분석 주제에 맞게 분석모델이 진행되는지에 대해 검증평가함

- 이러한 과정을 거쳐 개발된 분석 모델은 마지막으로 분석모델에 대한 매뉴얼 작성 및 확산을 위한 자산화를 수행하고 확산 과정을 통한 지속적 분석모델 개선을 수행함
- 또한, 통합문화이용권(문화누리카드) 빅데이터를 활용한 문화소외계층 문화향유 실태분석 및 개선방안 연구의 요구사항을 분석하고 정리하여 적용방안을 도출함

&lt;표 2-1&gt; 요구사항 분석 및 적용방안 도출

구분	연구요구사항	적용방안
1	1-1. 통합문화이용권(문화누리카드) 시스템 내 보유 자료(2014.1.1. ~ 2017.12.31.) 카드 결제사용 건 약 14,500천 건 및 관련 자료)의 수집 및 분석	· 통합문화이용권(문화누리카드) - 문화누리카드 데이터 사용(홈페이지 회원, 카드 사용정보, 사용건수 등 수집 및 분석)
	1-2. 인터넷 공간(트위터, 페이스북, 인스타그램, 포털 블로그/지식in, 커뮤니티, 뉴스 등)의 크롤링을 통한 통합문화이용권(문화누리카드) 관련 기초 자료 수집 및 변화 분석 - 통합문화이용권(문화누리카드) 고객지원센터 접수 VOC(전화, 문자) 수집 및 변화 분석	· K-ICT(한국정보화진흥원), 네이버 데이터 랩, 구글 트렌드 오픈소스 기반의 플랫폼 사용
	1-3. 통합문화이용권(문화누리카드) 수혜자의 최신 IT기기(태블릿, 휴대폰 등) 사용 현황 (인터넷, 문자, 결제 등) 수집 및 분석 최신 IT기기를 활용한 홍보 및 사용률 증대 방안 제시	· 통합문화이용권(문화누리카드)의 카드사용내역, 가맹 점 분포와 통신사 데이터, 일반카드 데이터 사용내역과의 비교를 통한 가맹점 분석

2	2-1. 수혜자의 사용분야(업종)와 사용행태 등 분석 및 빅데이터 활용 접근 방법 제시, 통합문화사용권(문화누리카드) 발급 후 미사용자에 대한 지역별/연령별/계층별 상세 원인 분석 및 사용촉진을 위한 서비스 개발 방안 제시	<ul style="list-style-type: none"> <li>· 문화누리카드 사용데이터, 일반카드(비교군) 수혜자와의 비교를 통한 행태 패턴 도출</li> <li>· 수혜자 행태의 특징을 분석하여 효율적 “수혜자 맞춤형 서비스” 설계</li> <li>· 연관분석의 기본알고리즘 적용을 통해 수혜자-가맹점, 수혜자 중심의 가맹점 사용 도출</li> <li>· 수혜자 행태를 예측함으로써 가맹점 공급(지역별/아이템별)계획과 유치활용가능</li> </ul>
	- 통합문화사용권(문화누리카드) 가맹점의 사용처/계절/연령/계층/사용요인/분야(도서, 공연, 영화 등) 사용금액 등 시스템 제공 자료 및 현장실사 조사 결과를 활용한 다각도 관점에서의 행태 분석	<ul style="list-style-type: none"> <li>· 가맹점 사용현황(매출액, 사용분야(업종), 지역별 특징)을 분석함으로써 가맹점 효용성 분석</li> <li>· 가맹점 사용분야(업종)별/매출액별 DB와 GIS DB연계를 통한 데이터 상관분석</li> <li>· 수혜자 불편 해소 및 가맹점 신규 유치 지역의 도출을 통한 계획 수립 지원</li> </ul>
	2-2. 지원액 증가에 따른 사용행태 변화 분석 및 지속적 증가 예상에 따른 수혜자 지출 예측(의견 수렴 및 분석 결과 활용) 제시	<ul style="list-style-type: none"> <li>· 지출의 규모와 흐름을 파악하기 위해 데이터 전처리, 데이터분석, 결론 도출 순으로 사업을 진행</li> <li>· 과거 사용행태 기반으로 지원액 증가에 따른 사용행태 변화 비교분석 및 예측</li> </ul>
3	3-1. 지역별 특징을 반영한 가맹점 확대 및 실수혜자 사용률 증대 방안 제언	<ul style="list-style-type: none"> <li>· 공간분석과 문화누리카드 사용 등의 융합 데이터셋 구축</li> <li>· 공간정보 기반으로 지역별 특징을 분류하고 Hotspot, Coldspot 지역 분석</li> </ul>
	3-2. 부정수급 방지를 위한 예방적 모니터링 알고리즘 적용 방안 제언	<ul style="list-style-type: none"> <li>· Anomaly Detection (이상치 분석) 알고리즘 모델링 적용을 통한 이상결제 방지</li> </ul>
	3-3. 예산 증액 등 수혜자 맞춤형 서비스 제공을 위한 의사결정 및 정책 수립 제언 - 향후 통합문화사용권(문화누리카드) 사업 활성화 방안 및 효과성 제고를 위한 빅데이터 활용방안 및 개선과제 제언	<ul style="list-style-type: none"> <li>· 카드발급대상자, 수혜자, 가맹점 기준으로 의사결정 모형 개발</li> <li>· 전문 빅데이터 TFT 팀을 통해 중·장기적 통합문화사용권(문화누리카드) 맞춤형 개선 과제 제언</li> </ul>

## 제2절 데이터 수집

## 1. 핵심 데이터의 이해

- 통합문화이용권(문화누리카드)은 저소득층의 삶의 질 향상과 문화격차 해소를 위한 문화 복지 사업으로 기초생활수급자, 차상위 계층에게 문화예술, 국내여행, 체육활동을 지원함

[그림 2-2] 문화누리카드 지원대상



- 통합문화이용권(문화누리카드) 주요 내용

- 문화예술진흥법 제 15조의 3, 제 15조의 4항을 근거로 문화체육관광부, 한국문화예술위원회, 전국 광역·기초 지방자치단체 및 17개 시도 지역 주관처가 추진함
- 문화누리카드는 기초생활수급자 및 차상위 계층(6세 이상)을 대상으로 문화예술·여행·체육 분야에 사용 가능한 ‘문화누리카드’발급 지원(1인당 연간 6만원)함

&lt;표 2-2&gt; 문화누리카드 주요 내용

구분	주요내용
목적	소외계층에게 문화예술·여행·체육 분야 향유 지원으로 삶의 질 향상 및 문화격차 해소
예산	총 992억 원(문예기금 699억 원, 지방비 293억 원)
기간	2017년 1월 ~ 12월
근거	문화예술진흥법 제 15조의3, 제15조의 4
추진경위	통합형 여가 바우처 도입, 저소득층의 문화향유기회 제공
사업주최	문화체육관광부, 한국문화예술위원회, 전국 광역·기초 지방자치단체
사업주관	17개 시도 지역 주관처

## － 문화누리카드 신청 자격

<표 2-3> 문화누리카드 신청 자격

구분		자격내용
기 초	수급자	<ul style="list-style-type: none"> <li>· 기초생계급여 수급자(기준 중위소득 30%)</li> <li>· 기초의료급여 수급자(기준 중위소득 40%)</li> <li>· 기초주거급여 수급자(기준 중위소득 43%)</li> <li>· 기초교육급여 수급자(기준 중위소득 50%)</li> </ul>
차 상 위	차상위 자활	· 수급자가 아닌 가구로 소득인정액이 기준중위소득의 50% 이하인 근로능력이 있는 미취업자
	장애아동수당, 장애수당	<ul style="list-style-type: none"> <li>· 장애수당은 국민기초생활보장법에 의한 수급자 및 차상위 계층의 18세 이상 등록 장애인 중 경증장애인에게 지급</li> <li>· 장애아동수당은 국민기초생활보장법에 의한 수급자 및 차상위 계층의 18세 미만 재가 장애아동에게 지급</li> </ul>
	차상위 본인부담경감	· 기준세대의 소득인정액이 기준중위소득의 50%이하로 희귀난치성 질환자, 만성질환자, 18세 미만의 아동 중 대상자가 부양의무자가 없거나 부양의무자가 있어도 부양능력이 없거나 부양을 받을 수 없는 자
	저소득 한부모가족	· 모 또는 부와 만 18세 미만(취학시 만 22세 미만)의 자녀로 이루어진 가정으로 소득인정액이 기준중위소득 52% 이하인 한부모 가족
	차상위 계층 확인서 발급	· 소득인정액이 기준중위소득의 50%이하이나 부양의무자 등으로 기초수급자, 차상위 대상이 되지 못하는 소득 계층
	장애인연금 '부가급여' 수급자 중 차상위 계층	<ul style="list-style-type: none"> <li>· 장애인연금은 소득인정액이 대통령령으로 정하는 금액 이하인 18세 이상의 중증장애인에게 지급</li> <li>· 장애인 부가급여 수급자 중 차상위 계층인 경우</li> </ul>

## □ 문화누리카드 사용분야(업종) 분석을 통한 특징 도출

- 주요 사용처인 문화예술, 여행, 체육 분야를 중심으로 사용분야(업종) 분포, 카드사용내역, 카드 사용 건수 등을 통해 수혜자 별 카드 사용 특징, 사용분야(업종)별 카드 사용 내역 등의 특징을 분석함

<표 2-4> 문화누리카드 주요 사용처

구 분	업 종
문화 예술	· 공연장, 미술관, 영화관, 서점, 헌책방, 만화방, 음반판매점, 악기소매점, 사진관, 수공예 문화상품, 화방, 정보화마을, 문화센터
여행	· 숙박 : 호텔, 콘도, 게스트이 · 운송수단 : 항공, 철도, 고속버스, 시외버스, 여객선, 렌트카 · 관광여행사 · 놀이공원(주요 테마파크, 워터파크 등) 및 스키장 · 지역축제 및 관광 명소(휴양림) · 온천(온천법 16조에 의거 시장·군수의 허가를 받은 온천에 한함)
체육	· 국내 4대(축구, 농구, 야구, 배구) 프로스포츠 경기관람 · 국내 4대 프로스포츠 구단 응원용품 (경기장 주변) · 국제스포츠경기대회(국내 개최) 및 기타 스포츠경기 관람·입장권 · 운동용품(체육사, 체육용품점에 한함) · 체육시설(수영장, 빙상장, 승마장, 볼링장, 탁구장, 당구장, 체력단련장 등)

## 2. 데이터 수집

### □ 통합문화이용권(문화누리카드) 시스템 내 보유 자료(2014.1.1. ~ 2017.12.31. 카드 결제사용 건 약 15,464천 건 및 관련 자료)의 수집

- 문화누리카드 데이터는 각 회원정보 파일에서 홈페이지 회원, 카드 회원, 문화누리카드 관리자(전국 3,400여개의 주민센터 담당자와 지역 담당자 등), 카드 사용 정보 등으로 구축함
- 데이터 수집 항목으로는 개인회원정보 파일의 경우 통합문화카드시스템의 회원정보로 ID, 비밀번호, 성명, 생년월일, 주소, 전화번호 등이 수집되며 카드 발급자 및 관련 정보에는 기본 개인회원정보에 카드번호 등이 추가됨

- － 문화누리카드 관리자의 경우 기관ID, 비밀번호, 성명, 주민등록번호(앞 7자리), 소속부서명 등이 수집됨

<표 2-5> 데이터 수집 테이블 목록

데이터 위치	DB 명	수집 데이터															
내부 데이터	통합문화사용권(문화누리카드) 시스템 보유 DB	카드 결제 사용 건 약 15,464천 건 및 관련자료															
	통합문화사용권(문화누리카드) 고객센터 DB	고객지원센터 접수 VOC(전화, 문자) 자료															
외부 데이터	카드사 데이터	<table><tr><td>블록_코드</td></tr><tr><td>행정동_코드</td></tr><tr><td>기준_연월</td></tr><tr><td>카드_업종_코드</td></tr><tr><td>카드_소상공인_업종_코드</td></tr><tr><td>사용분야(업종)_수</td></tr><tr><td>당월_문화누리카드_매출_총건수</td></tr><tr><td>당월_문화누리카드_매출_총금액</td></tr><tr><td>연간_매출_총건수</td></tr><tr><td>연간_매출_총금액</td></tr><tr><td>연간_환산_카드_매출_총금액</td></tr><tr><td>남성_매출_총건수</td></tr><tr><td>남성_매출_총금액</td></tr><tr><td>여성_매출_총건수</td></tr><tr><td>...</td></tr></table>	블록_코드	행정동_코드	기준_연월	카드_업종_코드	카드_소상공인_업종_코드	사용분야(업종)_수	당월_문화누리카드_매출_총건수	당월_문화누리카드_매출_총금액	연간_매출_총건수	연간_매출_총금액	연간_환산_카드_매출_총금액	남성_매출_총건수	남성_매출_총금액	여성_매출_총건수	...
	블록_코드																
행정동_코드																	
기준_연월																	
카드_업종_코드																	
카드_소상공인_업종_코드																	
사용분야(업종)_수																	
당월_문화누리카드_매출_총건수																	
당월_문화누리카드_매출_총금액																	
연간_매출_총건수																	
연간_매출_총금액																	
연간_환산_카드_매출_총금액																	
남성_매출_총건수																	
남성_매출_총금액																	
여성_매출_총건수																	
...																	
	공간정보 데이터	<table><tr><td>블록_코드</td></tr><tr><td>행정동_코드</td></tr><tr><td>행정구역선</td></tr><tr><td>건축물대장(정보)</td></tr><tr><td>도로명주소</td></tr><tr><td>인구밀도</td></tr><tr><td>...</td></tr></table>	블록_코드	행정동_코드	행정구역선	건축물대장(정보)	도로명주소	인구밀도	...								
블록_코드																	
행정동_코드																	
행정구역선																	
건축물대장(정보)																	
도로명주소																	
인구밀도																	
...																	

	공공데이터	건축물정보	
		주소	면적
		구조	용도
		층수	...
	통계청	인구총조사	
		나이	성별
		주소	자녀수
		...	

### － 통합문화이용권(문화누리카드) 데이터 목록

<표 2-6> 통합문화이용권(문화누리카드) 데이터

정보 구분		데이터 수집 항목	데이터 설명
회원 정보 파일	홈페이지 회원	회원 ID, 비밀번호, 성명, 영문성명, 주민/외국인등록번호, 주소, 전화번호, 휴대폰번호, 이메일 주소, 발급사유	문화누리 카드 시스템의 회원 정보
	카드 회원		
문화누리카드 발급자 및 관련정보		성명, 영문성명, 주민/외국인등록번호, 주소, 전화번호, 휴대전화번호, 이메일주소, 발급사유, 세대주 성명, 세대주 주민등록번호, 세대주 관계, 회원ID, 카드번호, 카드비밀번호, CVC번호, 만 14세 미만 아동의 개인정보 처리 시 법정대리인 정보(이름, 고유식별번호)	문화누리 카드 신청 및 발급, 사용정보
문화누리카드 관리자 (전국 3,400여개 주민센터 담당자, 지역 주관처 담당자 등)		기관ID, 비밀번호, 성명, 주민등록번호(앞7자리), 소속부서명, 전화번호, 휴대폰	문화누리 카드 시스템 관리자
카드 사용 정보		사용분야(업종) 정보, 결제일, 결제 금액	문화누리 카드 사용 건수

## 제3절 데이터 정제 및 구축

## 1. 데이터 정제 및 전처리

□ 전체 데이터를 정제하여 구축할 데이터는 다음과 같음

- 각 기관별로 수집한 데이터들을 하나의 융합 데이터 셋으로 만들기 위해 시군구를 key값으로 연계하여 데이터를 구축함

[그림 2-3] 문화누리카드 지원 대상

통합문화이용권 시스템 보유 DB				카드 발급자 및 관련 정보			
회원 ID	비밀번호	성명	영문성명	성명	영문성명	주민등록번호	외국인등록번호
주민등록번호	외국인등록번호	주소	전화번호	주소	전화번호	휴대전화번호	세대주 성명
휴대폰번호	이메일주소	발급사유	...	세대주 관계	회원ID	카드번호	...
원시자료				원시자료			
문화누리카드 관리자				카드 승인 정보			
기관 ID	비밀번호	성명		가맹점 정보	결제일	결제정보	
소속부서명	주민등록번호	전화번호					
읍면동 단위				원시자료			

□ 데이터 비식별 조치

- 개별적으로 관리되고 있는 데이터를 분석한 결과 개인정보가 사용되고 있음이 확인됨에 따라 데이터 정제 단계에서 데이터 비식별 조치를 진행하였고, 비식별 조치 단계는 비식별 조치 기법 ‘식별자 조치 기준’과 ‘속성자 조치 기준’으로 분리하여 진행함
- 식별자 조치를 통한 비식별 조치 기법 적용
  - 식별자란, 개인 또는 개인과 관련된 사물에 고유하게 부여된 값 또는 이름을 칭함
  - 정보 집합물에 포함된 식별자는 원칙적으로 삭제 조치하게 되어있으나 데이터 사용 목적상 반드시 필요한 식별자는 비식별 조치 후 활용함

&lt;표 2-7&gt; 식별자 예시

<ul style="list-style-type: none"> <li>· 고유 식별정보(주민등록번호, 여권번호, 외국인 등록번호, 운전면허번호)</li> <li>· 성명(한자영문 성명, 필명 등 포함)</li> <li>· 상세주소(구 단위 미만까지 포함된 주소)</li> <li>· 날짜정보 : 생일(양/음력), 기념일(결혼, 돌, 환갑 등), 자격증 취득일 등</li> <li>· 전화번호(휴대전화번호, 집전화, 회사전화, 팩스번호)</li> <li>· 의료기록번호, 건강보험번호, 복지 수급자 번호</li> <li>· 통장계좌번호, 신용카드 번호</li> </ul>	<ul style="list-style-type: none"> <li>· 각종 자격증 및 면허 번호</li> <li>· 자동차 번호, 각종 기기의 등록번호 &amp; 일련번호</li> <li>· 사진(정지사진, 동영상, CCTV 영상 등)</li> <li>· 신체 식별정보(지문, 음성, 홍채 등)</li> <li>· 이메일주소, IP 주소, Mac 주소, 홈페이지 URL 등</li> <li>· 식별코드(ID, 사원번호, 고객번호 등)</li> <li>· 기타 유일 식별번호 : 군번, 개인사업자의 사업자등록번호 등</li> </ul>
--	--

#### — 속성자 조치 기준을 통한 비식별 조치 기법 적용

- 속성자란, 개인과 관련된 정보로서 다른 정보와 쉽게 결합하는 경우 특정 개인을 알아볼 수도 있는 정보를 칭하며, 정보 집합물에 포함된 속성자도 데이터 사용 목적과 관련이 없는 경우에는 원칙적으로 삭제함
- 데이터 사용 목적과 관련이 있는 속성자 중 식별요소가 있는 경우에는 가명처리, 총계 처리 등의 기법을 활용하여 비식별 조치를 진행함. 희귀병명, 희귀경력 등의 속성자는 구체적인 상황에 따라 개인 식별 가능성이 매우 높으므로 엄격한 비식별 조치가 필요함

&lt;표 2-8&gt; 속성자 예시

특성	속성자
개인 특성	<ul style="list-style-type: none"> <li>· 성별, 연령(나이), 국적, 고향, 시·군·구명, 우편번호, 병역여부, 결혼여부, 종교, 취미, 동호회·클럽 등</li> <li>· 흡연 여부, 음주 여부, 채식 여부, 관심 사항 등</li> </ul>
신체 특성	<ul style="list-style-type: none"> <li>· 혈액형, 신장, 몸무게, 허리둘레, 혈압, 눈동자 색깔 등</li> <li>· 신체검사 결과, 장애유형, 장애등급 등</li> <li>· 병명, 상병코드, 투약코드, 진료내역 등</li> </ul>
신용 특성	<ul style="list-style-type: none"> <li>· 세금 납부액, 신용등급, 기부금 등</li> <li>· 건강보험료 납부액, 소득분위, 의료 급여자 등</li> </ul>
경력 특성	<ul style="list-style-type: none"> <li>· 학교명, 학과명, 학년, 성적, 학력 등</li> <li>· 경력, 직업, 직종, 직장명, 부서명, 직급, 전 직장명 등</li> </ul>
전자적 특성	<ul style="list-style-type: none"> <li>· 쿠키정보, 접속일시, 방문일시, 서비스 사용 기록, 접속로그 등</li> <li>· 인터넷 접속기록, 휴대전화 사용기록, GPS 데이터 등</li> </ul>
가족 특성	<ul style="list-style-type: none"> <li>· 배우자·자녀·부모·형제 등 가족 정보, 법정대리인 정보 등</li> </ul>

### － 비식별 조치 방법

- 본 사업에는 가명처리, 총계처리, 데이터 삭제, 데이터 범주화, 데이터 마스킹 등 여러 가지 기법을 복합적으로 활용하여 비식별 조치를 진행하였음
- 이는 ‘가명 처리’기법만 단독으로 활용한 경우 충분한 비식별 조치로 보기 어렵기 때문임

<표 2-9> 비식별 조치 방법

처리기법	예시	세부기술
가명처리 (Pseudonymization)	홍길동, 35세, 서울 거주, 한국대 재학 → 임격정, 30대, 서울 거주, 국제대 재학	① 휴리스틱 가명화 ② 암호화 ③ 교환 방법
총계처리 (Aggregation)	임격정 180cm, 홍길동 170cm, 이콩쥐 160cm, 김팔쥐 150cm → 물리학과 학생 키 합 : 660cm, 평균키 165cm	④ 총계처리 ⑤ 부분총계 ⑥ 라운딩 ⑦ 재배열
데이터 삭제 (Data reduction)	주민등록번호 901206-1234567 → 90년대 생, 남자  개인과 관련된 날짜정보(합격일 등)는 연단위로 처리	⑧ 식별자 삭제 ⑨ 식별자 부분삭제 ⑩ 레코드 삭제 ⑪ 식별요소 전부삭제
데이터 범주화 (Data Suppression)	홍길동, 35세 → 홍씨, 30~40세	⑫ 감추기 ⑬ 랜덤 라운딩 ⑭ 범위 방법 ⑮ 제어 라운딩
데이터 마스킹 (Data Masking)	홍길동, 35세, 서울 거주, 한국대 재학 → 홍○○, 35세, 서울 거주, ○○대학 재학	⑯ 임의 잡음 추가 ⑰ 공백과 대체

- － 각각의 기법에는 이를 구현할 수 있는 다양한 세부기술이 있으며, 데이터 사용 목적과 기법별 장단점 등을 고려하여 적절한 기법, 세부 기술을 선택 및 활용하여 개인정보 비식별 조치를 진행함

- 원본데이터에 주민등록번호가 존재한다면 주민등록번호를 삭제하거나, 앞자리만 남기고 별표(\*)를 활용하여 마스킹 처리를 진행함. 또한, 날짜의 경우 년도를 제외하고 마스킹 처리를 진행하며, 연령의 경우 평균연령으로 통일하여 데이터 마스킹 처리를 진행함

[그림 2-4] 데이터 삭제 및 마스킹



## □ 데이터 정제 및 전처리

- 제공 받은 데이터들은 파일이 각각 나뉘어져 있어 데이터들 간의 연관관계를 분석하기 어려운 문제점이 발견됨
- 또한 수작업으로 수집되고 비식별 조치된 원시 데이터는 오기입, 누락, 중복 기입, 이상치 등을 포함하고 있기 때문에 보다 정확한 데이터 분석을 위해 정제 및 전처리를 진행함

&lt;표 2-10&gt; 비식별 조치 방법

방식	설명
데이터 여과 (Filtering)	오류 발견, 보정, 삭제 및 중복성 확인 등의 과정을 통해 데이터 품질을 향상시키는 기술
데이터 변환 (Transformation)	데이터 유형 변환 등 데이터 분석이 용이한 형태로 변환하는 기술 정규화(normalization), 집합(aggregation), 계층 생성 등의 방법 활용
데이터 정제 (Cleansing)	데이터 분석이 용이하도록 유사 데이터 및 연계가 필요한 데이터(또는 DB)들을 통합하는 기술
데이터 축소 (Reduction)	분석 컴퓨팅 시간을 단축할 수 있도록 데이터 분석에 활용되지 않는 항목 등을 제거하는 기술

- 데이터 전처리 시 데이터 필터링, 데이터 변환, 데이터 정제에 대한 고려사항을 <표 2-11>와 같이 정의하여 통합 데이터를 정리·진행함

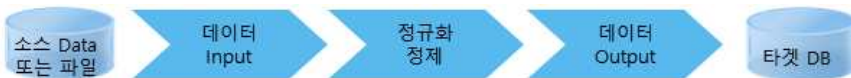
&lt;표 2-11&gt; 데이터 전처리 시 고려사항

항목	고려사항
데이터 필터링	<ul style="list-style-type: none"> <li>· 데이터 필터링 기준 정의</li> <li>· 데이터 처리 전후 생성된 데이터 중복 확인</li> <li>· 유의미한 데이터 선별을 위한 사전 정의된 필터일 기준과 비교 검증</li> <li>· 수집된 데이터 품질 기준 부합 여부 및 오류 등 관리자에게 피드백</li> <li>· 필터링 기준에 의해 처리된 데이터의 이력 제공</li> </ul>
데이터 변환	<ul style="list-style-type: none"> <li>· 수집 데이터의 유형 분류에 대한 기준 적용</li> <li>· 데이터 변환에 대한 함수 및 변환 구조 정의</li> <li>· 데이터 변환 시 변환 형식에 준하여 변환이 이루어 졌는지 확인</li> </ul>
데이터 정제	<ul style="list-style-type: none"> <li>· 정제 유형 사전 정의 및 속성 값 부여</li> <li>· 데이터 유형별 정제 방법 사전 정의된 형태로 자동 수행</li> <li>· 결측치, 잡음 데이터 처리</li> <li>· 데이터 불일치성 (단위, 표현형식, 코드체계 등) 교정</li> </ul>

#### - 오류 정제 실행 프로세스

- 데이터 전처리 프로세스는 수집된 소스데이터를 입력 받아 여과, 변환, 정제, 축소 등의 정규화 과정을 거쳐 타겟 DB로 출력함

[그림 2-5] 프로세스 진행도



#### - 데이터 구조화

- 수집된 데이터 중 구조화가 안 된 데이터가 있을 경우, 구조화된 데이터로 변형을 하여야만 그 다음 처리가 가능함
- 레코드에 필드명이 오류가 난 상태로 있을 때 레코드를 분리하여 변환 함수를 사용해 코드를 변환함
- 필드 명을 자동으로 입력시키고, 필드 값 중 한글 깨짐 현상을 변환 함수로 사용하여 한글코드로 변환시켜 구조화를 진행함

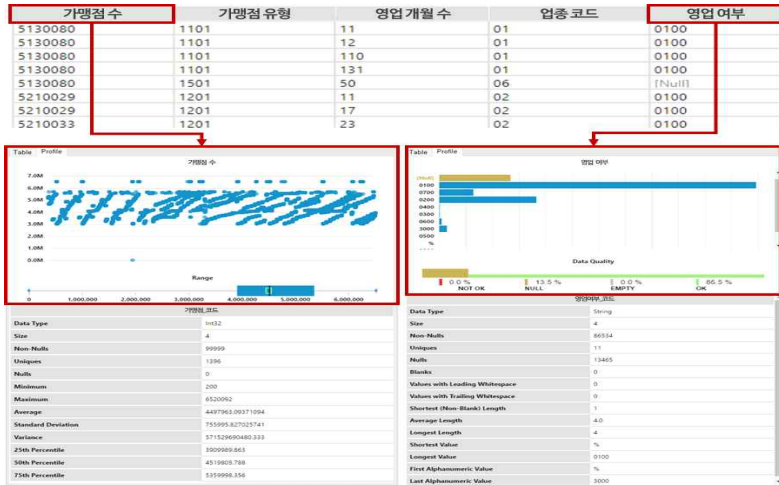
[그림 2-6] 데이터 구조화를 위한 입력데이터처리 결과 데이터

입력데이터					처리 결과 데이터				
47290-100016169	2010	5130080	1101	11	47290-100016169	2010	5130080	1101	11
47290-100016170	2010	5130080	1101	12	47290-100016170	2010	5130080	1101	12
47290-100017791	2010	5130080	1101	110	47290-100017791	2010	5130080	1101	110
47290-100018093	2010	5130080	1101	131	47290-100018093	2010	5130080	1101	131
47290-100020930	2010	5130080	1501	50	47290-100020930	2010	5130080	1501	50
47840-100012572	2010	5210029	1201	11	47840-100012572	2010	5210029	1201	11
47840-100012711	2010	5210029	1201	17					

## – 각 필드 값 데이터 프로파일링 및 품질확인

- 구조화된 데이터가 출력되면 전체 데이터의 품질을 파악하기 위하여 각 필드별 데이터 프로파일링을 실행함
- 각 필드별 값에 대한 데이터 타입, 길이, Null 값 허용 여부 등 규칙을 파악하고 전체 데이터의 품질을 확인함

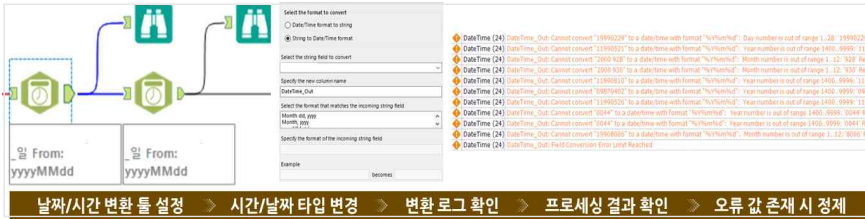
[그림 2-7] 프로파일링 화면



## – 시간날짜 타입 필드 오류 확인

- 시간날짜는 데이터 분석의 필수 항목이므로, 날짜별·기간별 추이나 합계, 평균 등을 분석해야 할 경우 날짜 타입 또는 잘못 입력된 값이 존재하면 분석에 오류를 일으킬 수 있음
- 이런 오류를 방지하기 위해 사전에 시간날짜 타입의 필드 오류를 확인하는 작업을 진행함

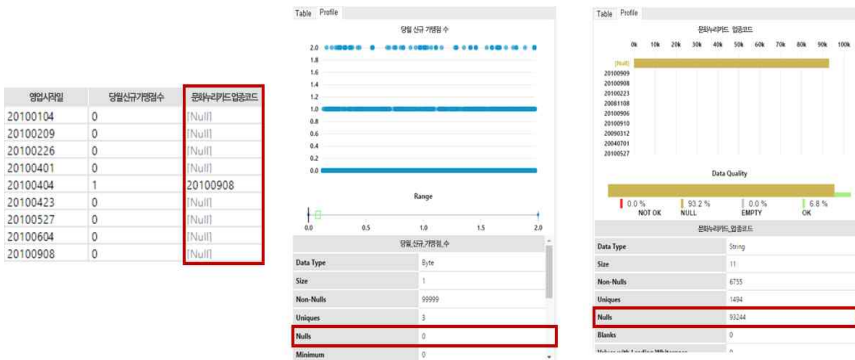
[그림 2-8] 필드 오류 확인 절차



### - Null 값에 대한 비즈니스 규칙 확인

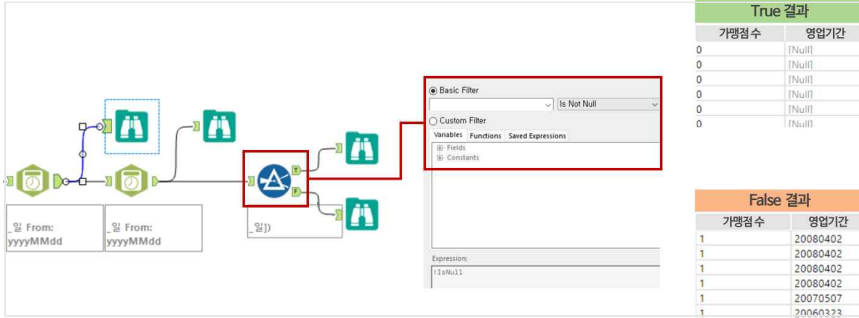
- Null 값이 있는 필드는 필드에 대한 비즈니스 규칙을 파악하여 Null값 허용 여부를 파악하고, 이후 Null값을 어떻게 처리 할지를 정의함
- 예로, ‘사용분야(업종) 수’ 필드는 Null값이 존재하면 안 되며, ‘사용분야(업종) 가입’의 값이 ‘0’이면 ‘사용분야(업종) 미가입’의 값은 Null, ‘1’이면 사용분야(업종) 가입에 들어있어야 하는 비즈니스 규칙을 세우는 것임

[그림 2-9] 비즈니스 규칙 확인



- 위 테이블 ‘당월 신규 사용분야(업종) 수’ 필드 값은 오른쪽 프로파일링 결과 Null값이 0인 것으로 확인되어 데이터 품질이 좋은 것으로 판단함
- ‘문화누리카드 업종코드’ 필드는 프로파일링 결과 Null값이 93,244개인 것으로 확인되어 ‘문화누리카드 업종코드’ 필드에 대한 데이터 품질을 검증함

[그림 2-10] Null 규칙 검증 화면



- 위와 같은 데이터 전처리 프로세스를 통해 통합문화이용권(문화누리카드) 시스템 내 보유한 카드 사용 건(15,464,584 건, '14.1.1 ~ '17.12.31)의 분석을 위한 정제 DB를 구축함
- 주요 필드를 중심으로 개인정보 관련 정보를 제거<sup>2)</sup>하여 데이터 분석이 가능한 DB형태로 구성하였고, 제공 DB필드는 22개로 정제 및 Null값을 제외하면 실제 사용 건은 약 15,000천 건의 분석 DB로 구성되어있음

&lt;표 2-12&gt; 정제된 필드 목록

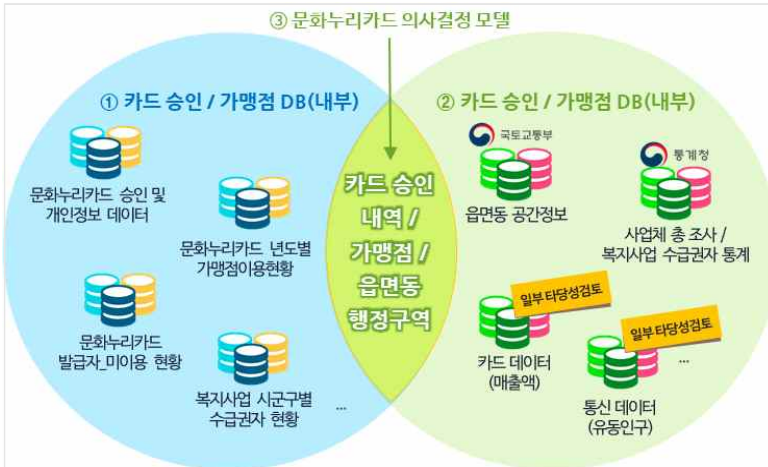
사용번호	사용일시	취소일자	수혜자_출생년도	수혜자_성별
발급구분	재충전여부	카드구분	수혜자_소속_자치구	수혜자_소속_주민센터
수혜자_발급자격	카드_대리발급여부	사용분야(업종) 장르 코드	사용분야(업종)명	사용분야(업종) 사업자등록 번호
사용분야(업종) 주소 지자치구명	사용금액	보조금 사용액	개인충전금 사용액	보조금 잔액
개인충전금 잔액	결제장소_온/오프라인구분	-	-	-

2) 카드 사용번호, 주소 블록화(구 단위), 개인정보(이름, 주민번호 등) 제거하여 제공받아 활용

## 2. 데이터 구축

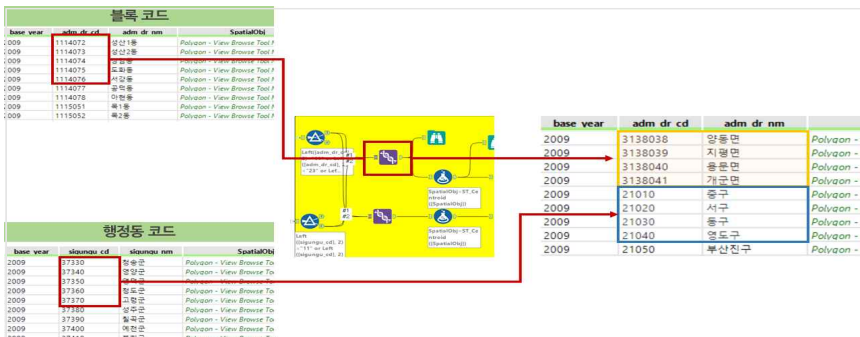
□ 데이터 분석 시 다양한 데이터를 융합하여 의미 있는 인사이트 도출을 진행함

[그림 2-11] 통합문화이용권(문화누리카드) 의사결정 모델을 위한 융합 데이터셋 구성(안)



- 데이터 융합 시 테이블 조인, 다중 조인, 병합, 필드 어펜드 등 다양한 방법으로 분석을 진행함
  - 예를 들면, 블록 코드와 행정동 코드가 동일한 스키마 구조이므로 병합을 통하여 조각된 데이터를 하나의 데이터로 통합함

[그림 2-12] 데이터 융합 예시 (1)



- 등록일시와 문화누리카드 사용분야(업종) 코드가 융합된 테이블과 문화누리카드 소상공인 사용분야(업종) 코드의 다른 스키마 구조이고 조인 필드가 동일한 것이 있을 때 테이블 조인을 통하여 하나의 데이터로 통합함

[그림 2-13] 데이터 융합 예시 (2)

- 이상치(이상점, Outlier)란, 관측된 데이터와 범위에서 많이 벗어난 작은 값이나 큰 값을 의미하며, 어떤 의사결정을 하는데 필요한 데이터를 분석할 경우 이렇게 이상한 값들에 의해 의사결정에 영향을 미칠 수 있으므로 제거함
- 이상치 판별 알고리즘은 사분위수, 정규분포, 앤드류스 그림, 마하라노비스 거리, 표준정규분포 등이 있으나 가장 많이 적용하고 있는 표준정규분포 (Z스코어) 방법으로 이상치를 제거함

[그림 2-14] 표준정규분포 :  $Z=(x-\text{평균})/\text{표준편차}$ 

Z	Z의 두번									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9825	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
4.0	0.999968314									
4.5	0.999996599									
5.0	0.999999713									
5.5	0.999999981									
6.0	0.999999999									

- Z스코어 방식의 공식은  $Z=(X-\text{평균})/\text{표준편차}$ 이며, Z스코어 방식으로 이상치를 제거하는 과정의 예제를 정의함
- Z스코어 범위 설정, 테스트 필드 설정, 그룹바이 필드 설정을 진행하며, ‘남성\_매출\_총건 수’ 필드에서 설정한 Z스코어 범위를 벗어난 데이터 값은 True(이상치)이고 범위를 넘기지 않은 데이터 값은 False(정상치)로 확인됨

[그림 2-15] 테스트 수행 결과

입력 데이터 테스트 필드			테스트 필드의 이상치 수행 결과 확인			
AVG 03TMS	AVG 04TMS	AVG 05TMS	GroupBY	StdDev	Avg	Outlier
0.1	0.11	0.17	39020620100051000001	0.028841	0.02015	True
0.1	0.11	0.17	39020620100051000001	0.028841	0.02015	True
0.01	0.01	0.01	39020620100051000001	0.028841	0.02015	False
0.01	0.01	0.01	39020620100051000001	0.028841	0.02015	False
0.1	0.11	0.17	39020620100051000001	0.028841	0.02015	True
0.1	0.11	0.18	39020620100051000001	0.028841	0.02015	True
0.01	0.01	0.01	39020620100051000001	0.028841	0.02015	False

- 데이터 정제(Data Cleaning)를 통한 원천 데이터 정확도 검증 결과는 아래와 같음

[그림 2-16] 통합된 데이터 정리화면

필드	포문 그래프	속성	최소값	최대값	합금	표준 편차	영도	고유값수	유치수
📊 불록코드		연속형	1.000	27962.000	13981.500	8072.078	0.000	—	27962
📊 행정동코드		범주형	—	—	—	—	—	169	27962
📊 거로_년월		범주형	—	—	—	—	—	—	27962
📊 년간_매출_총금액		연속형	0.000	32510.000	15383.140	8433.133	0.022	—	27962
📊 소상공인_업종_코드		범주형	—	—	—	—	—	—	27962
📊 연례대_10_매출_총건수		연속형	1.000	582.000	50.139	59.551	2.901	—	27962
📊 년간_완산_매출_총금액		범주형	—	—	—	—	—	—	4141
📊 법인_매출_총건수		연속형	101.000	101.000	101.000	0.000	—	—	2
📊 법인_매출_총건수		범주형	—	—	—	—	—	2	27962
📊 개인_매출_총건수		연속형	0.000	83.000	2.909	5.679	3.304	—	27962
📊 개인_매출_총금액		범주형	—	—	—	—	—	2	27962
📊 남성_매출_총건수		연속형	1.000	33050.000	15488.390	9025.853	0.054	—	27962
📊 남성_매출_총금액		범주형	—	—	—	—	—	26	27962
📊 여성_매출_총건수		연속형	0.000	41.000	3.174	3.127	4.266	—	27962
📊 여성_매출_총금액		범주형	—	—	—	—	—	1	27962

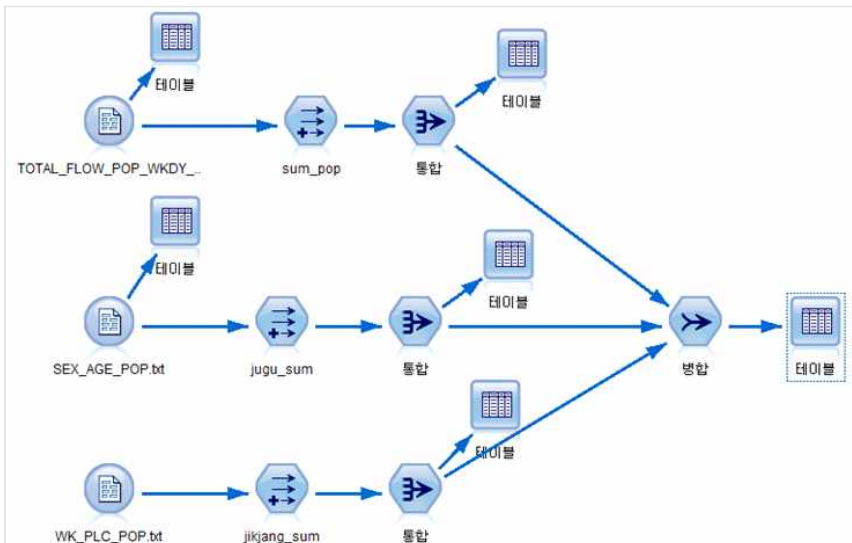
[illegible]

- 위와 같은 데이터 구축 방법을 통해 통합문화이용권(문화누리카드) 시스템 내 카드 사용 건 15,464,584건과 유동인구(통신사), 카드사, 용도지역의 공간데이터를 융합한 데이터셋을 구축함

[그림 2-17] 구축에 활용한 카드 데이터 일부지역

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	162620	
161	12626	2940300000	4	48	185	3	1	58	2	1	47900000	95	7	0.00	25	0	1	11700000									
162	12640	2880000000	4	48	185	3	1	58	3	1	80000000	95	6	0.00	12	1	1	11700000									
163	13408	0	3	48	138	7	1	62	3	1	13000000	130	9	1.62	12	0	1	13000000									
164	1368	3070000000	4	52	165	3	1	11	1	60	80000000	95	6	0.00	40	0	1	80000000									
165	137	134000000	4	48	85	0	2	40	2	1	12	42000000	95	6	4.35	25	0	1	42000000								
166	2121	357000000	4	33	87	3	2	48	4	1	12	50000000	100	4	6.67	35	0	1	50000000								
167	13854	0	4	62	67	4	1	53	1	1	11	30000000	130	3	0.00	24	0	1	30000000								
168	12200	0	4	70	129	3	1	45	3	1	11	13000000	90	9	0.00	23	0	1	13000000								
169	1308	1797000000	4	48	47	1	1	12	3	1	30	12000000	130	3	0.00	38	0	1	12000000								
170	21146	1364000000	4	62	185	3	1	44	1	1	11	13700000	90	3	0.00	35	0	1	13700000								
171	17883	3842000000	4	69	207	4	1	44	1	1	12	42000000	95	4	6.88	36	0	1	42000000								
172	21040	1740000000	4	30	109	3	1	28	2	0.9	42000000	85	1	0.00	12	0	1	42000000									
173	3812	788000000	3	20	125	3	1	48	3	1	11	39000000	95	9	11.12	23	0	1	39000000								
174	472	494000000	4	40	246	4	1	15	4	0.00	120	187	4	4.67	11	0	1	12000000									
175	124	0	4	62	294	0	1	55	1	1	14	13000000	130	4	0.00	59	0	1	13000000								
176	11711	0	4	62	252	4	1	44	1	1	12	42000000	85	4	7.62	23	0	1	42000000								
177	21579	0	4	52	112	4	1	48	1	1	5	42000000	85	4	0.00	23	0	1	42000000								
178	7120	0	4	28	113	0	1	80	2	1	100	1000000	130	5	0.00	59	0	1	10000000								
179	12427	0	48	96	7	1	1	100	3	1	1	10000000	130	7	0.00	12	0	1	10000000								
180	16370	0	4	48	62	1	1	43	1	1	1	20000000	100	2	1.50	12	0	1	20000000								
181	14908	0	4	41	130	1	1	49	3	1	5	20000000	100	5	0.00	24	0	1	20000000								
182	15902	618000000	4	48	112	7	2	25	2	0.9	20000000	100	7	0.00	24	0	1	4000000									
183	15903	618000000	4	48	112	7	2	28	3	1	20000000	130	6	0.00	12	0	1	4000000									
184	14907	0	48	112	0	1	1	49	4	2	618	44	7	0.75	12	0	1	4000000									
185	17084	3684000000	4	17	243	7	1	64	1	1	5	80000000	95	3	0.00	35	0	1	95000000								
186	17421	916000000	4	54	154	3	1	52	1	1	11	42000000	85	4	0.00	12	0	1	42000000								
187	21488	2188000000	3	42	367	10	1	88	1	1	17000000	85	4	0.00	24	0	1	17000000									
188	727	0	4	62	70	3	1	58	3	1	13000000	130	4	4.95	12	0	1	13000000									
189	4634	1891400000	4	48	142	3	1	36	1	1	30	11000000	85	5	0.00	12	0	1	11000000								
190	4638	430000000	4	75	107	2	1	48	3	1	1	10000000	130	8	0.00	59	0	1	10000000								
191	4638	430000000	4	75	107	3	1	47	3	1	13	20000000	130	7	0.00	47	0	1	20000000								
192	21222	494000000	3	22	179	20	1	43	1	1	1	95000000	85	7	0.00	35	0	1	95000000								

[그림 2-18] 융합 데이터셋 구축(예\_일부지역 대상 유동인구 DB)



## 제3장

---

### 데이터 분석 및 알고리즘 개발

제1절 : 주제별 표준분석모델

1. 주제별 표준분석모델 정의

제2절 : 주제별 데이터 분석

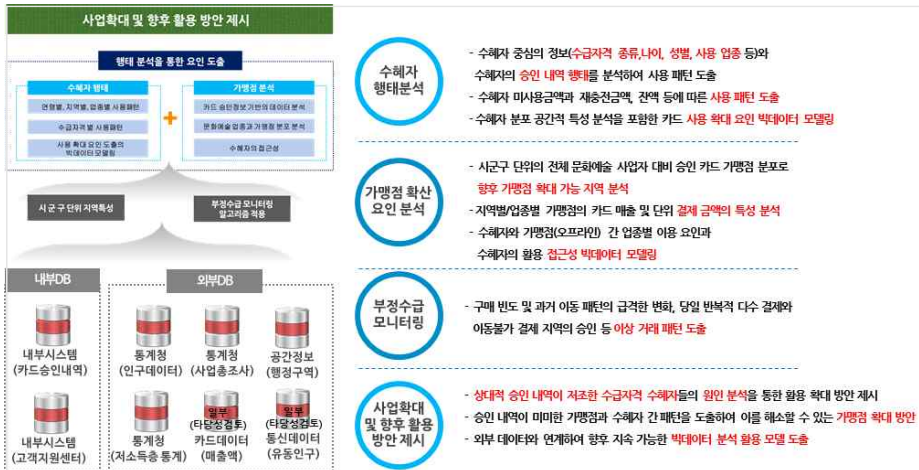
1. 분석 시나리오 설정
2. 주제별 데이터 분석

## 제1절 주제별 표준분석모델

### 1. 주제별 표준분석모델 정의

- 데이터기반 통합문화이용권(문화누리카드) 사업 변화의 내외부 환경요인과 수혜자 유형을 분석하여 수혜자에게 적합한 통합문화이용권(문화누리카드) 사업과 관련 정보를 제공하고, 문화향유에 특화된 산업의 규모와 흐름을 파악하여 패턴을 분석하고 문화 복지 개선방안을 도출하는데 활용할 수 있는 모델을 개발함
- 즉, 주제별 요인 범주에 따른 특성들을 수혜자 행태분석, 사용분야(업종) 확산 요인분석, 부정수급 모니터링, 사업 확대 및 향후 활용방안 4가지 측면으로 접근하였으며, 이에 따라 통합문화이용권(문화누리카드)의 사용패턴, 사용분야(업종) 확대 등 의사결정에 활용할 수 있는 정책 수립을 제안하는 것을 목표로 함

[그림 3-1] 주제별 의사결정 표준분석모델 개념 정의



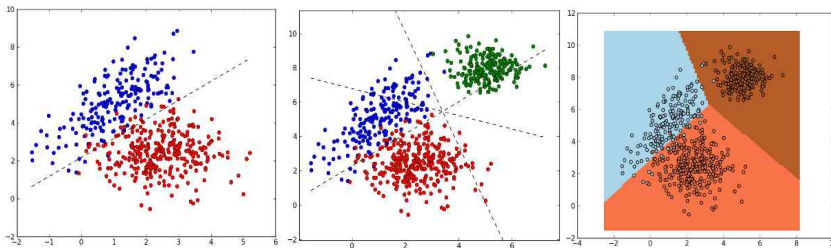
- 표준분석모델의 알고리즘을 구축하기 위한 주요 적용 알고리즘은 분류분석, 클러스터링, 이상치 탐지, 차원 축소, 연관성분석, 기계학습, 랜덤 포레스트, 그래디언트 부스팅 머신, 딥 러닝 등이 있음

## □ 주요 적용 알고리즘

### － 분류분석(Classification Analysis)

- 데이터의 실체가 어떤 그룹에 속하는지 예측하는데 사용하는 데이터 마이닝 기법으로 고객 기록을 특정한 등급으로 나눈다는 점에서 클러스터링과 유사하지만, 분류분석은 군집분석과 달리 각 계급이 어떻게 정의되는지 미리 알아야 할 필요가 있음
- 분류(Classification)는 어떤 객체가 불량인지 우량인지 또는 생존하느냐 못하느냐와 같이 0과 1로 구분하는데 활용되거나, ABCD 중에서 어느 범주에 속하는지 또는 1:2:3:4:5등급 중에서 어느 등급에 속하는지 등과 같이 객체를 정해놓은 범주로 분류하는 목적이 있음
- 분류를 위해 사용되는 데이터마이닝 기법은 K-최근접 이웃(K-Nearest Neighborhood), 의사결정 나무(Decision Tree), 베이저안 정리(Bayesian Theorem)를 사용한 분류, 인공신경망 (Artificial Neural Network), 지지벡터기계(Support Vector Machine) 등이 있음

[그림 3-2] Linear Methods for Classification

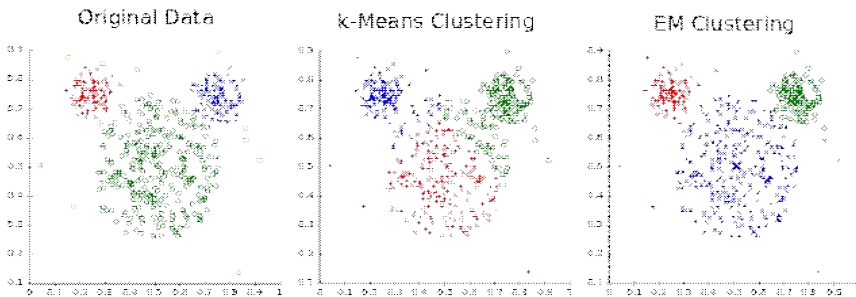


- 분석기법을 활용하여 통합문화이용권(문화누리카드) 시스템 내 보유자료 분석, 인터넷 공간의 통합문화이용권(문화누리카드) 관련 기초자료 분석, 통합문화이용권(문화누리카드) 고객지원센터 접수 VOC 변화 분석, 통합문화이용권(문화누리카드) 수혜자 최신 IT 기기 사용현황 분석, 수혜자의 사용분야(업종)와 사용형태 분석을 진행함

### － 클러스터링(Clustering)

- 기체를 다수의 매트릭스에서 상호 유사한 세그먼트 또는 클러스터로 그룹화하는 기법으로 고객 세분화가 클러스터링의 실제 예이며, 다양한 클러스터링 알고리즘 중에서 가장 널리 사용되는 것이 K-평균(K-means)임

[그림 3-3] Clustering Analysis

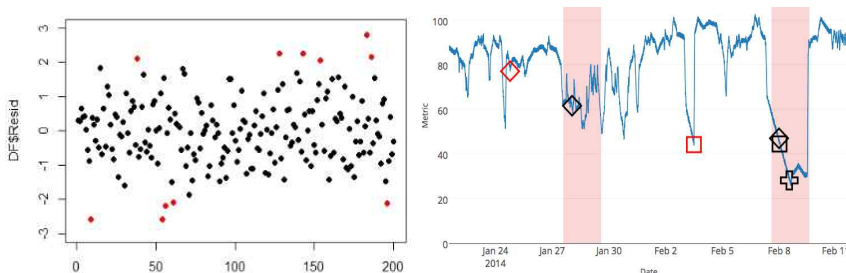


- 분석 기법을 활용하여 통합문화이용권(문화누리카드) 발급 후 미사용자에 대한 지역별연령별계층별 상세 원인 분석을 진행함

### － 이상치 탐지(Anomaly detection)

- 예상치 못한 이벤트 또는 결과를 식별하는 프로세스, 보안, 사기 등의 분야에서는 모든 거래를 철저히 조사하기란 불가능하므로 가장 비 일반적인 거래에 체계적으로 플래그를 지정해야 함
- 분석 기법을 활용하여 이상결제에 대한 패턴 분석 알고리즘 개발을 진행함

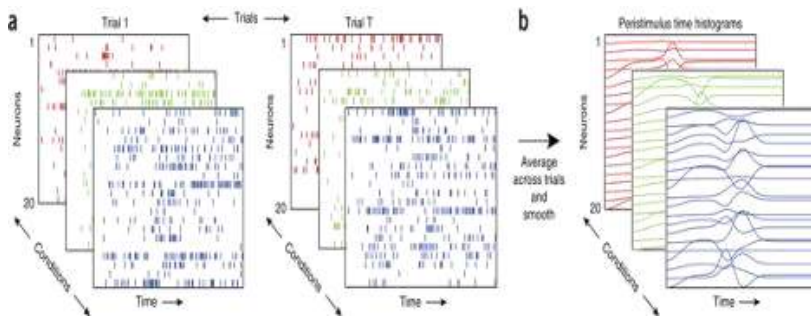
[그림 3-4] Anomaly detection



### — 차원 축소(Dimension Reduction)

- 고려 대상 변수의 수를 줄이는 프로세스로 조직이 더 많은 데이터를 캡처할수록 예측에 사용 가능한 예측 변수(또는 특징)의 수도 급격히 증가함
- 특정 문제에 대해 가치 있는 정보를 제공하는 데이터를 식별하는 것만 해도 상당한 작업이 소요됨
- 주성분 요소 분석(Principal components analysis, PCA)은 일련의 원시 특징을 평가해 이를 상호 독립적인 인덱스로 축소함

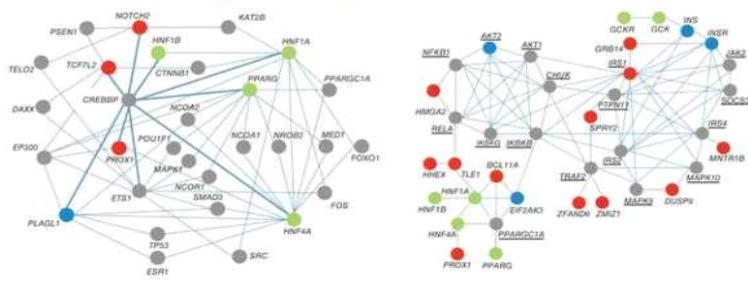
[그림 3-5] Dimension reduction



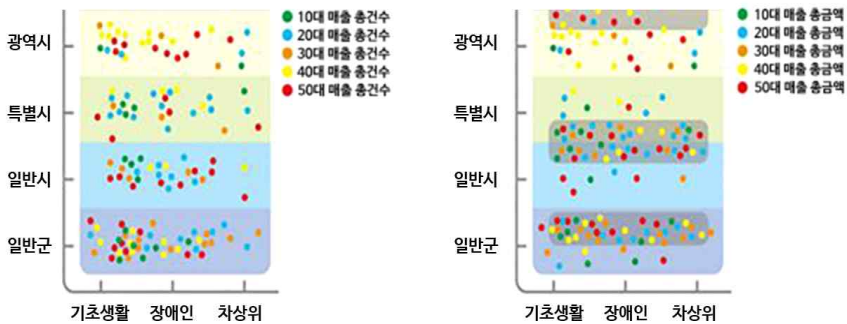
### — 연관성분석(Association Analysis)

- 흔히 장바구니분석(Market Basket Analysis), 서열분석(Sequence Analysis)이라 불림
- 예를 들어, 대형 할인점의 계산대에서 매일 수집되는 구매 트랜잭션 데이터는 유니크한 트랜잭션 ID로 구별되며, 주어진 고객이 구매한 아이템 항목의 집합으로 이루어짐
- 이 거대한 데이터에서 유용한 정보를 얻기 위해 연관 분석을 활용할 수 있음

[그림 3-6] Association Analysis



[그림 3-7] Sample Association Analysis

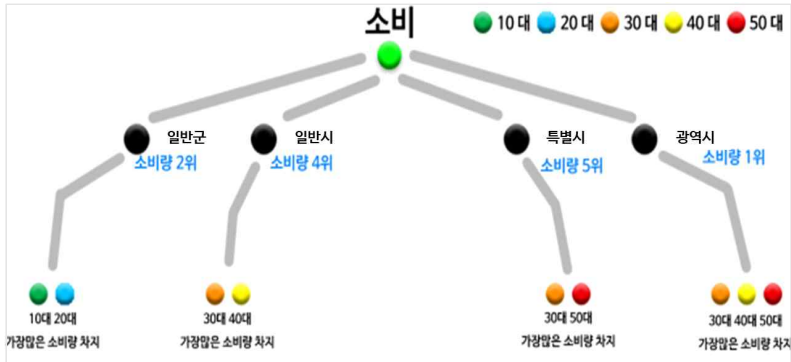


- 분석 기법을 활용하여 통합문화이용권(문화누리카드) 사용분야(업종)의 사용처·계절·연령·계층·사용요인·분야·사용금액 등 다각도 관점에서 형태를 분석하고, 지역별 특징을 반영한 사용분야(업종) 확대 및 실수혜자 사용을 증대 방안을 도출함

## — 기계학습

- 컴퓨터가 데이터를 기반으로 학습하고 예측하고, 성능을 향상시키는 알고리즘을 연구하고 구축하는 기술을 말하며, 가장 광범위하게 사용되는 지도 러닝 기법은 ‘의사결정 트리(Decision trees)’로 모 집단을 대상 변수에 대해 동질적인 더 작은 조각으로 점진적으로 분할하는 규칙 집합을 학습하는 자율 학습 방법을 사용함

[그림 3-8] Decision Tree를 사용한 소비 예측

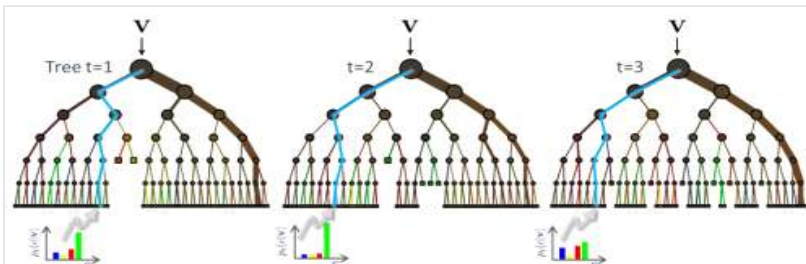


- 분석 기법을 활용하여 지원액 증가에 따른 사용형태 변화 분석, 지속적 증가 예상에 따른 수혜자 지출 예측, 예산 증액 등 수혜자 맞춤형 서비스 제공을 위한 의사결정 및 정책을 수립함

#### — 랜덤 포레스트(Random forests)

- 널리 사용되는 총체적 학습 방법으로, 다수의 의사결정 트리를 학습한 다음 트리 전반에 걸친 평균을 구해 예측을 산출함. 이 평균 프로세스는 일반화 가능한 솔루션을 제공하며 데이터의 불규칙 잡음(Random noise)을 걸러내는 효과가 있음

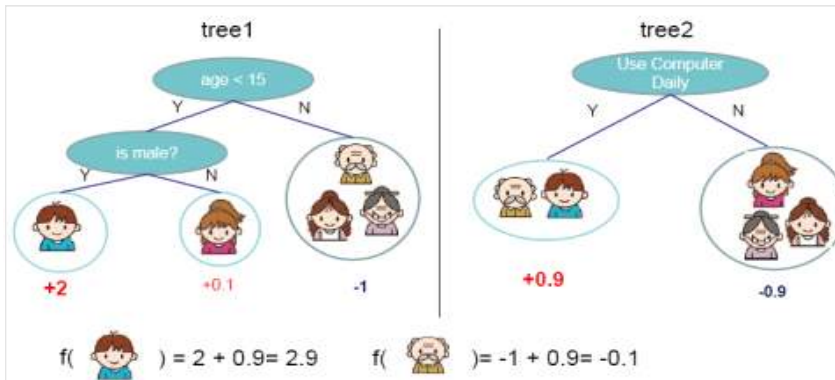
[그림 3-9] Random forests



### - 그래디언스 부스팅 머신(Gradient boosting machine, GBM)

- 의사결정 트리의 시퀀스 교육을 통해 예측 모델을 생성하는 방법으로, 연속되는 트리가 이전 트리의 예측 오류를 수정해 나감

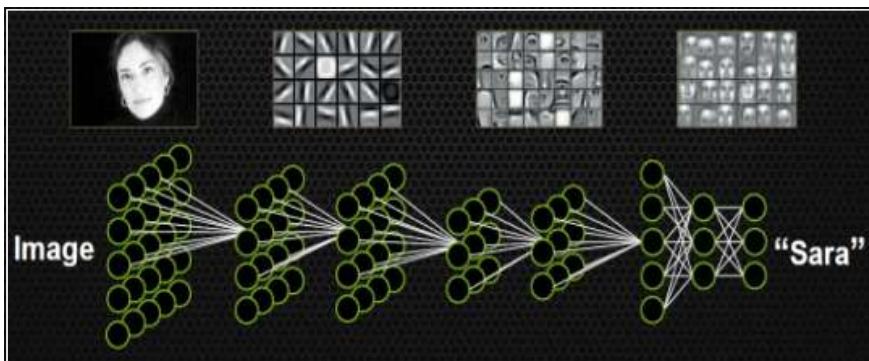
[그림 3-10] Gradient boosting machine



### - 딥 러닝(Deep learning)

- 데이터의 고수준 패턴을 복잡한 다계층 네트워크로 모델링하는 방법으로 문제를 모델링하는 가장 일반적인 방법이며, 머신러닝의 가장 어려운 문제를 해결할 잠재력을 지니고 있음

[그림 3-11] Deep learning



## 제2절 주제별 데이터 분석

## 1. 분석 시나리오 설정

□ 도출된 주제에 대한 분석 과정의 흐름을 사전에 정의하여 효율적인 분석이 이루어질 수 있도록 시나리오를 정의함

## □ 주제별 분석 시나리오

## - 수혜자 형태 분석

- 수혜자 특성에 따른 사용내역과 미사용금액, 재충전금액, 잔액과의 상관관계 및 분류 패턴 분석을 진행함. 수혜자별 카드사용내역은 연령별·지역별·사용분야(업종)·수급자격 등 총 11개 속성으로 구성되어 있음
- 생활 반경 및 사용분야(업종) 분포를 감안하여 구축된 시군구 단위 융합 데이터셋을 활용하여 수혜자 카드사용금액 합을 목표 값으로 빅데이터 분석 모델링을 통해 사용 형태를 분석함. 또한, 시군구별 카드사용금액 추정 값에 따른 변수 중요도를 분석하여 해당 요인을 도출함

## - 사용분야(업종) 확산 요인 분석

- 사용분야(업종) 확대 가능 지역의 카드사용 건수 및 금액을 세분화하여, 유동인구(통신사 DB) 및 상업지역(국토교통부 DB)과 사용분야(업종)(특정지역 대상) 간 관계 패턴을 분석함
- 분석 결과를 활용하여 향후 사용분야(업종) 확대 가능 지역 분석을 수행함
- 또한, 전국 읍면동 단위로 수혜자의 카드 발급 위치정보<sup>3)</sup>와 사용분야(업종) 간의 거리를 변수로 설정하여, 사용내역 등 추가적인 빅데이터 분석이 필요함

## - 이상결제내역 패턴 분석

- 수혜자별 사용 빈도 및 이동 패턴의 급격한 변화, 당일 반복적 다수 결제와 이동불가 결제 지역의 사용 등 이상거래 패턴을 도출함

3) 현재 개인정보보호법 보호로 인해 수혜자의 카드 발급 위치정보를 파악 할 수는 없지만, 향후 개인정보보호를 코드화 처리를 통해 데이터를 수집 할 수 있다면, 사용분야(업종)와 거리, 위치정보 등 상관관계 분석을 통해 사용분야(업종)별 가맹점의 추가 등록 등에 대한 정보를 파악 할 수 있음

## □ 사업확대 및 향후 활용 방안 제시를 위한 분석 시나리오

### - 수급 자격별 수혜자 분석을 통한 확산

- 수혜자 수급 자격에 따라 상대적으로 사용 내역이 저조한 자격 구분의 원인을 분석함
- 이를 통해 장애를 가진 수급자가 문화예술 공연 관람시 이동할 때 겪게 되는 교통문제를 해결해주는 정책 아이디어를 도출함

### - 사용 한도와 사용 건수 간 패턴 분석

- 지원한도와 사용금액과 카드 사용 건수 간 분석을 통해 통합문화누리사용권의 한도 상향 효과(5만원('16년) -> 6만원('17년))의 객관적 결과를 도출함
- 카드 사용 후 재충전 시 본인 충당금의 2배와 같은 상한액 제시 등 새로운 한도 향상 방안을 제시하는 정책 아이디어를 도출함

### - 사용분야(업종)와 사용 건수의 지역 분석

- 카드사용 건수와 사용분야(업종) 간 점포 수 차이가 높은 지역의 사용분야(업종) 활동을 강화하는 등 정책 아이디어를 도출함

## 2. 주제별 데이터 분석

### □ 수혜자 분석

#### － 수혜자 탐색적 분석

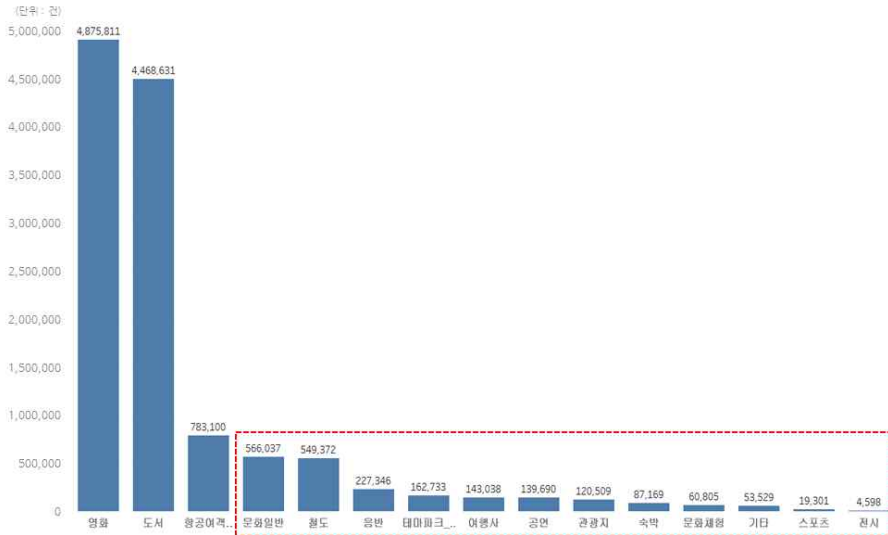
- 다음은 연령대별 카드 사용내역(사용건수)을 기반으로 카드를 많이 사용한 연령대 살펴보면, 20~30세대 사용 비중이 가장 높았으며, 그 중 98년생~01년생이 높았음. 그 다음으로는 40대 후반(1971년생~1968년생)이 높았으며, 65세 이상은 1942생이 높은 것으로 나타남(그림 3-12 참고)

[그림 3-12] 연령대별 카드 사용내역



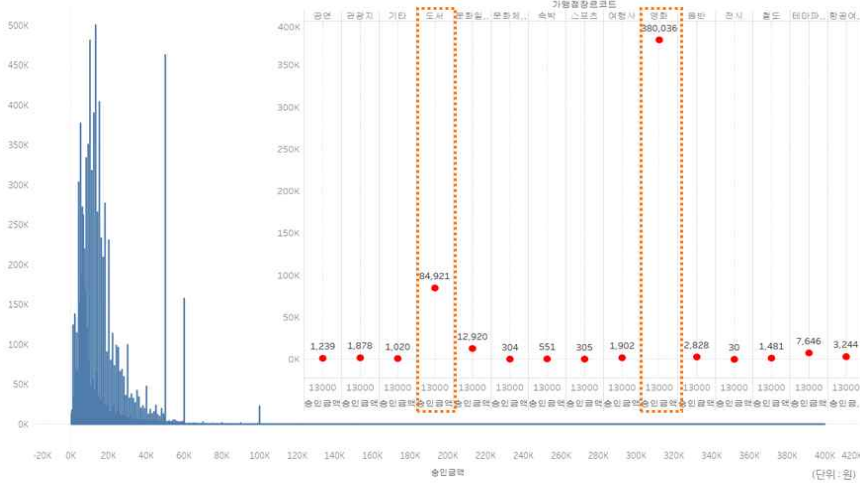
- 수혜자의 사용분야(업종)별 카드 사용내역을 분석한 결과(그림3-13)를 보면 영화, 도서에 주로 편중되어 있지만, 이외에도 항공여객, 문화일반, 철도 순으로 교통과 문화 분야에 많이 사용한 것으로 나타남

[그림 3-13] 사용분야(업종)별 수혜자의 카드사용



- 금액에 따른 사용분야(업종) 선택 패턴 분석을 위해 수혜자의 카드사용 금액과 사용분야(업종) 내역간의 분석을 진행하였음
- 가장 많이 사용한 카드 사용금액은 13,000원으로 나타났으며, 이를 사용분야(업종)의 종류로 상세히 살펴보면, 영화(380,036건), 도서(84,921건) 순이고, 이는 1회 사용당 단가가 저렴한 영화나 도서의 사용액이 많은 것을 나타냄
- 그림 3-14와 같이 사용금액을 살펴보면 2만원까지의 사용금액이 전체 약 75%정도 차지하고 있음. 이는 1회 사용당 사용금액이 많은 2만원까지의 저비용 문화서비스 발굴이 필요하며, 문화예술체육 분야의 세부적인 사업으로 사용분야(업종) 확대를 모색하는 계획이 요구됨

[그림 3-14] 수혜자의 카드사용 금액과 사용 사용분야(업종) 내역



### — 빅데이터 분석 방법론을 적용한 수혜자 분석

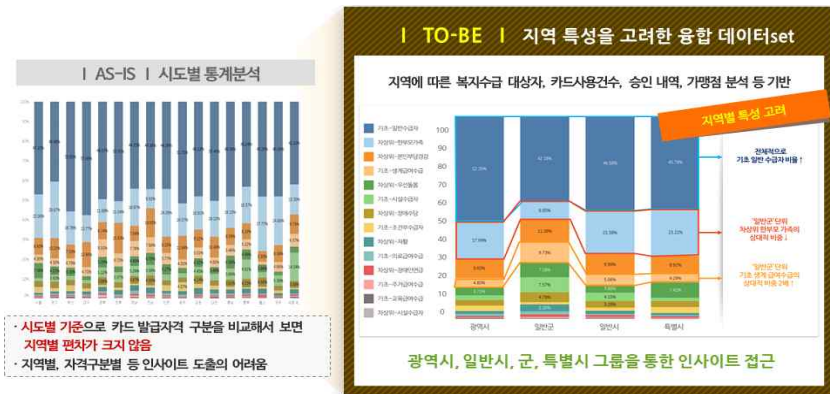
- 기존의 데이터 분석 방법론으로는 시도별 집계 데이터 기반의 분석 결과물이 도출되어 단순 통계에 가까운 형태로 분석하였음
- 본 연구에서는 클러스터링 기법을 포함하는 빅데이터 분석 방법론을 적용하여 군집 인사이트를 도출하였음

[그림 3-15] 분석용 데이터셋 구축 및 빅데이터 분석 방법론



- 기존 시도별 분석 결과는 지역 간의 상호편차를 구분하기 어려웠지만, 연구에서는 시도별 분석 내용을 ‘특별시’, ‘광역시’, ‘일반시’, ‘일반군’ 지역으로 구분하고 분석을 진행하여 유사한 지자체 패턴을 도출하고 수혜자 특성을 파악하였음
- 즉 전국 지자체(시군구)를 클러스터링(Clustering)기법을 적용하여 분석을 진행하였는데 광역시는 부산, 대구, 인천, 광주, 대전, 울산 6개의 구분으로 나누었으며, 일반시는 세종시를 포함한 자치시, 창원, 수원, 용인, 고양, 성남 등 총 77개의 시로, 일반군은 도 및 광역시의 하부 행정구역이자 기초지방자치단체로써 총 82곳의 군으로, 특별시는 서울특별시 1개의 구분으로 분류하였음
- 특별시, 광역시, 일반군, 일반시로 구분하여 클러스터링 분석을 진행함으로써 지역별 특성이 고려된 분석 결과를 얻을 수 있었음

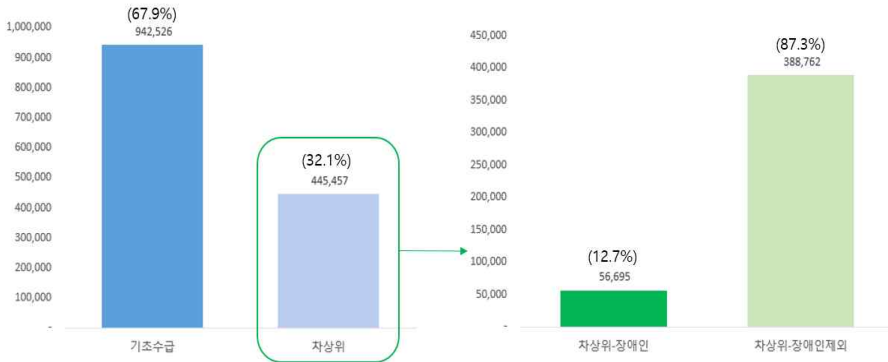
[그림 3-16] 지역별 특성 도출



- 위와 같은 분석 방법론을 적용하여 계층그룹별 수혜자 분석을 통한 사용내역 특징을 분석하고자 함
- 카드 사용내역을 기준으로 ‘계층그룹별 수혜자 분포’와 ‘연령대별 수혜자 분포’, ‘장애인 계층의 사용분야(업종) 이용현황 분석’을 살펴보고자 함
  - 그림 3-17 을 통해 2017년 카드사용에 따른 계층그룹별 수혜자 분포를 분석한 결과 수혜자 분포는 기초수급자 계층이 67.9%(942,526건)를 차지하고, 차상위 계층이 32.1%(445,457건)를 차지하는 것으로 나타남

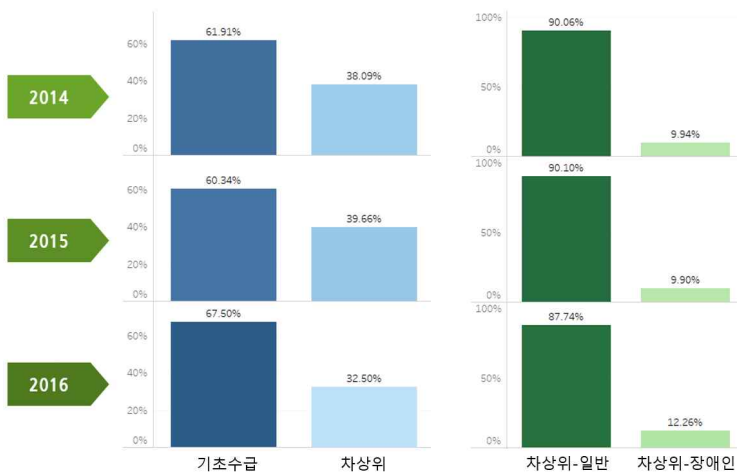
- 차상위 계층 중 장애인의 분포는 12.7%(56,695건), 장애인이 아닌 분포는 87.3%(388,762건)로 나타남

[그림 3-17] 계층그룹별 수혜자 분포 (2017)



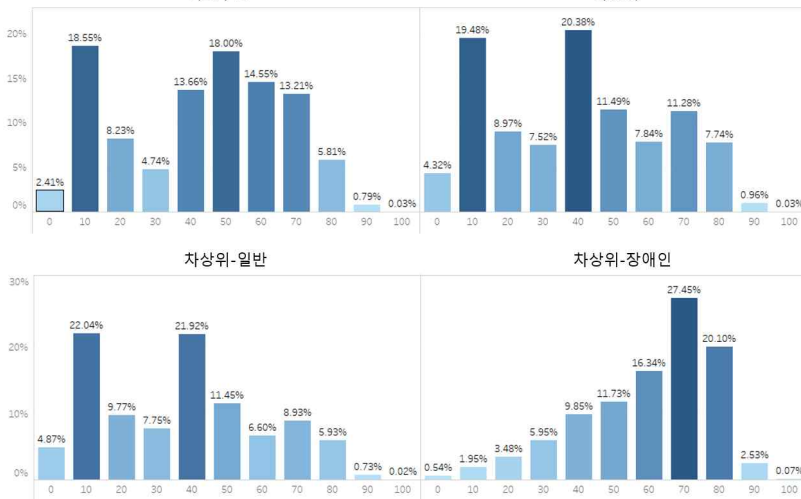
- 분류기준은 카드사용내역 데이터의 자격구분 필드를 활용함. 장애인은 차상위-장애수당, 차상위-장애인연금에 해당하는 항목을 말함
- 그림 3-18을 통해 '14년 ~ '16년까지 계층그룹별 수혜자 분포를 살펴보았을 때 '16년에는 '15년 대비 차상위 계층에서 장애인 계층이 약 2.3%증가한 것을 알 수 있음

[그림 3-18] 계층그룹별 수혜자 분포 (2014 ~ 2016)



- 카드사용에 따른 계층그룹의 연령대별 수혜자 분포를 살펴보면, 기초수급 계층에는 10대(18.6%)와 50대(18%), 차상위 계층은 10대(19.5%)와 40대(20.4%)가 주로 분포되어 있음
- 또한 차상위 계층 중 일반인과 장애인을 살펴보면, 차상위-일반인의 경우 10대(22.0%), 40대(21.9%), 차상위-장애인은 70대(27.5%), 80대(20.1%)가 주로 분포되어 있음

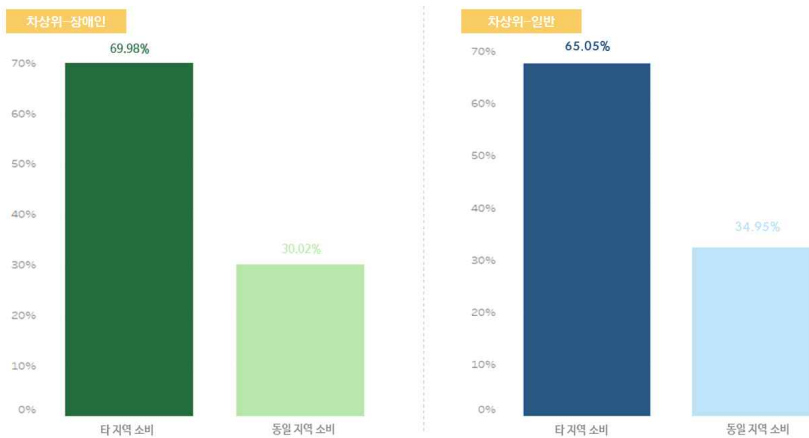
[그림 3-19] 계층그룹 연령대별 수혜자 분포  
기초수급 차상위



- 차상위-장애인은 70대와 80대가 많이 분포되어 있어, 다른 계층과 별도로 담당자의 방문 발급을 용이하게 하여 카드 발급을 활성화하고, 기초수급권자와 차상위 계층에서는 수혜자가 많이 분포되어 있는 40대 이상에 대해 통합문화이용권(문화누리카드)에 대한 홍보 및 적극적 지원이 필요함
- 정책적으로 연령대에 맞는 사용분야(업종)을 확대하고, 사용내역에 맞는 접근성을 강화시켜야 함
- 다음은 차상위 계층에서 장애인과 일반의 소비행태를 비교해서, 차상위-장애인의 경우 카드를 발급 받은 본인의 동네(본인 자치구)안에서 소비를 많이 하는가를 살펴보기 위해 "차상위-장애인"과 "차상위-일반"과 비교 분석하였음

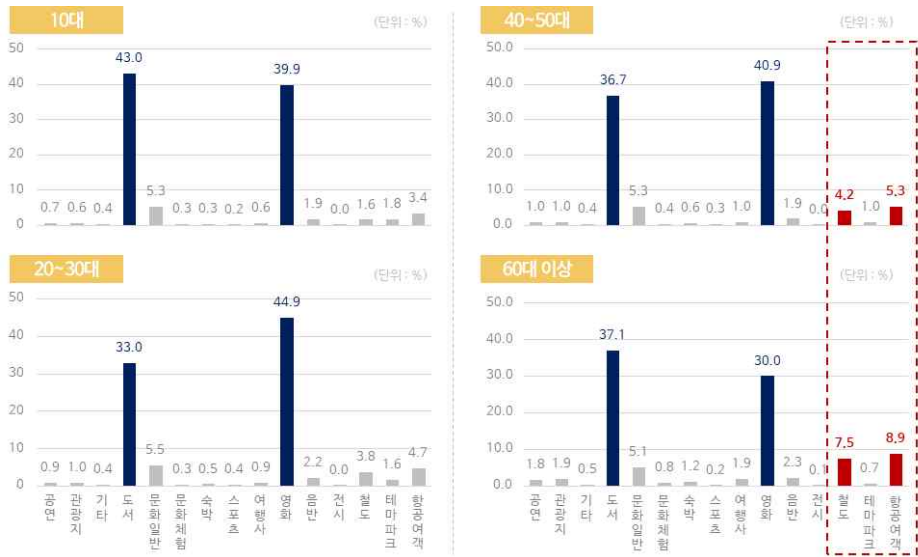
- 그 결과 그림 3-20을 살펴보면 차상위-장애인 계층은 본인 카드발급 지역(본인 거주지) 이외에 타 지역 소비가 69.9%로 동일 지역 소비(30.0%)보다 많은 것을 알 수 있었으며, 차상위-일반 역시 타 지역 소비가 65.0%로 동일 지역 소비(34.9%)보다 약 30% 이상 많았음

[그림 3-20] 장애인 계층 사용분야(업종) 이용현황 분석



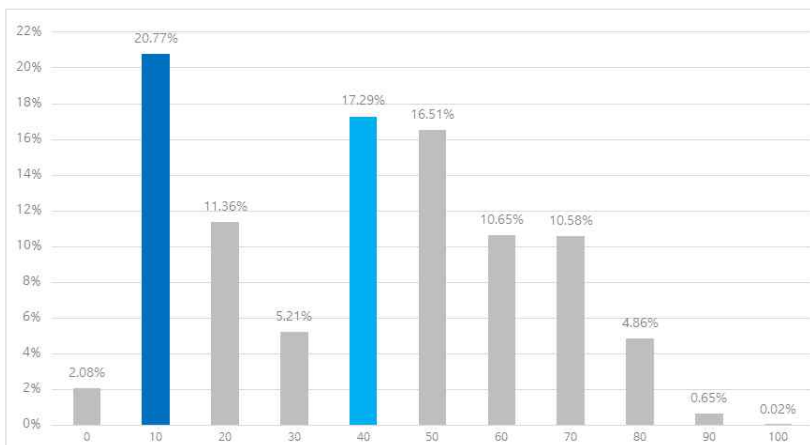
- 이는 거주지 근처의 사용분야(업종)를 한정적으로 사용하기 보다는 교통시설 등을 이용하여 타 지역에서 사용하는 것을 알 수 있음. 즉 타 지역 문화시설을 이용하기 위한 교통수단에 대해 확대 지원이 필요함
- 수혜자의 연령대별 사용분야(업종) 사용 선호도를 분석한 결과 그림 3-24를 통해 10대 사용 선호도를 살펴보면 도서가 42.9%로 가장 높았고 다음으로 영화가 39.9%를 보임
- 20~30대 사용 선호도는 영화가 44.8%, 도서가 32.9% 순으로 나타났고 40~50대의 경우 영화가 40.8%로 가장 높았으며 도서(36.74%)가 그 다음으로 20~30대와 유사한 분포를 보였으며, 60대 이상의 경우 도서가 37.09%로 가장 높았고 영화(30.0%)가 그 다음으로 순으로 나타남
- 모든 연령대에서 도서와 영화에 통합문화이용권(문화누리카드) 사용이 집중되고 있음

[그림 3-21] 사용 선호도 분석



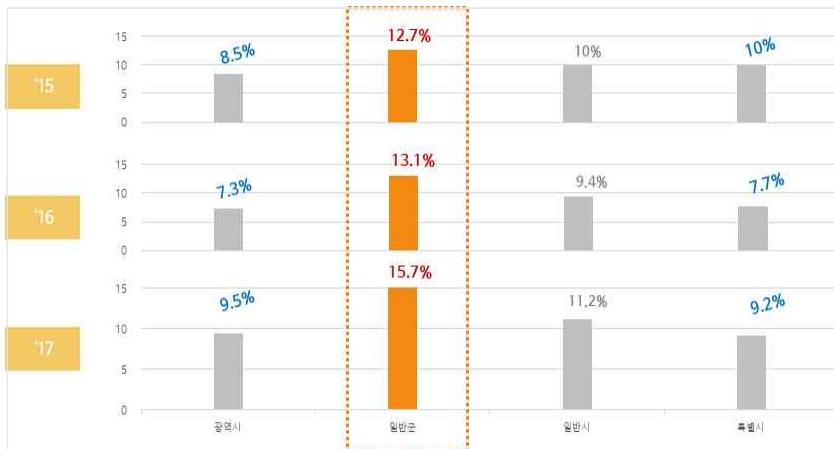
- 카드 수혜자를 연령대별로 분류해보면 10대가 20.77%로 가장 높고 40대가 17.29%, 50대가 16.51% 순으로 10대의 카드 사용 비중이 높은 것으로 나타났다 (그림 3-22 참고)

[그림 3-22] 전체 연령대 분포



- 40대~50대와 60대 이상의 경우 10대, 20~30대에 비해 항공여객과 철도의 이용 비율이 높게 나타나는데, 항공여객의 경우 40~50대 5.2%, 60대 이상 8.8%를, 철도의 경우 40~50대 4.19%, 60대 이상 7.5%를 나타냄
- 이는 고령자장애인의 경우 철도·항공여객 등 생활편의 수요가 늘어남에 따라 교통과 연계한 사용분야(업종) 지원이 추가적으로 필요함을 알 수 있음
- 연도별 발급카드 대비 미사용 카드 수를 살펴보면, 일반군의 경우 12.7%('15년)에서 15.7%('17년)로 약 3%정도 증가하였고, 특별시의 경우는 10%('15년)에서 9.2%('17년)로 약 8%정도 감소하였음
- 이는 일반군의 경우 특별시의 비해 카드는 발급하였지만, 사용되지 않은 카드 수가 많음을 의미하며, 이를 통해 일반군에 홍보, 사용분야(업종) 수의 증가 등 사용 확대를 위한 지원이 필요하다는 것을 알 수 있음
- 발급 카드 수 대비 미사용 카드 수, 즉 미사용 카드 수가 적고, 카드사용이 잘 된 지역을 '15년, '16년, '17년을 기준으로 상위 20곳의 지자체를 도출하면 표 3-1, 3-2, 3-3과 같음

[그림 3-23] 발급 카드 대비 미사용 카드 수 비율



&lt;표 3-1&gt; '15년 기준 사용카드 수 가 많은 주민센터 (상위 20위)

시군구	주민센터	미사용카드수 / 카드발급수
전북 ○○시	○○ 1 동 주민센터	0.1%
울산 ○구	○○동주민센터	0.2%
경남 ○○군	○○읍사무소	0.2%
제주 ○○시	○○ 2 동주민센터	0.3%
전북 ○○시	○○읍사무소	0.3%
전남 ○○시	○○동주민센터	0.3%
경북 ○○시	○○동주민센터	0.3%
전남 ○○시	○○ 2 동주민센터	0.3%
경북 ○○시	○○동주민센터	0.4%
전남 ○○시	○○동주민센터	0.4%
전남 ○○시	○○동주민센터	0.4%
충남 ○○시	○○동주민센터	0.4%
충남 ○○시	○○ 2 동 주민센터	0.4%
전북 ○○시	○○동주민센터	0.4%
울산 ○구	○○ 2 동주민센터	0.4%
전남 ○○시	○○면주민센터	0.5%
경남 ○○시	○○동주민자치센터	0.5%
전북 ○○시	○○동주민센터	0.5%
경북 ○○시	○○읍사무소	0.5%
전북 ○○시	○○면사무소	0.5%

&lt;표 3-2&gt; '16년 기준 사용카드 수 가 많은 주민센터 (상위 20위)

시군구	주민센터	미사용카드수 / 카드발급수
전북 ○○시	○○동주민센터	0.2%
경남 ○○시	○○동주민센터	0.2%
울산 ○구	○○동행복복지센터	0.2%
경북 ○○시	○○동주민센터	0.3%
전남 ○○시	○○동주민센터	0.3%
전북 ○○시	○○동주민센터	0.3%
경기 ○○시	○○ 3 동주민센터	0.3%
경북 ○○시	○○읍사무소	0.3%
전북 ○○군	○○면사무소	0.3%
경북 ○○시	○○동주민센터	0.4%
전북 ○○시	○○면사무소	0.4%
전북 ○○시	○○동주민센터	0.4%
전북 ○○군	○○읍 주민자치센터	0.4%
전북 ○○시	○○면사무소	0.4%
경기 ○○시	○○면사무소	0.4%
경남 ○○시	○○주민센터	0.4%
경북 ○○시	○○동주민센터	0.4%
전북 ○○군	○○면사무소	0.4%
전북 ○○시	○○면사무소	0.4%
대전 ○구	○○동주민센터	0.4%

&lt;표 3-3&gt; '17년 기준 사용카드 수 가 많은 주민센터 (상위 20위)

시군구	주민센터	미사용카드수 / 카드발급수
전남 ○○군	○○읍사무소	0.4%
울산 ○구	○○○○○동주민센터	0.5%
전북 ○○시	○○동 주민센터	0.5%
경기 ○○시	○○동주민센터	0.5%
울산 ○구	○○ 2 동주민센터	0.5%
전북 ○○군	○○면사무소	0.5%
경남 ○○군	○○면사무소	0.5%
울산 ○구	○○ 1 동행복복지센터	0.6%
울산 ○구	○○ 3 동주민센터	0.6%
경남 ○○군	○○면사무소	0.6%
충남 ○○군	○○면주민센터	0.6%
전북 ○○시	○○면사무소	0.7%
경북 ○○시	○○동주민센터	0.7%
전남 ○○시	○○면 주민센터	0.7%
경북 ○○군	○○면사무소	0.7%
경북 ○○군	○○면사무소	0.8%
경남 ○○군	○○면사무소	0.8%
충남 ○○시	○○면 행정복지센터	0.8%
전북 ○○군	○○면사무소	0.8%
경북 ○○시	○○면사무소	0.8%

- 카드 사용의 특징을 분석하기 위해 활발하게 사용하는 두 지역을 선정하였으며, '16년도 전북 ○○시의 ○○동 주민센터와 '17년도 울산 ○구의 ○○2동 주민센터의 발급 카드 수 대비 미사용 카드 수를 비교하였음
- 전북 ○○시 ○○동 주민센터의 경우 카드사용 평균 나이가 51.3세, 사용금액이 13,262,640원, 미사용자 나이 84세, 미사용카드 수 1개, 발급카드 수 259개, 총 사용금액/ 카드발급수(원/수) 51,010원으로 도출됨. 총 259개 카드발급수의 사용카드수가 258개로 타 지자체에 비해 매우 높은 사용률을 보이며, 1건에 대한 미사용자 나이는 84세임
- 울산 ○구 ○○2동 주민센터의 경우 카드사용 평균나이가 41.2세, 사용금액이 12,824,830원, 미사용자 나이 22세, 미사용카드 수 1개, 카드발급 수 200개, 총 사용금액/ 카드발급 수(원/수) 63,805원으로 도출됨. 총 200개

카드발급수의 미사용카드 수가 1개이며 사용카드 수가 199개로 매우 높은 사용률을 보이며, 1건에 대한 미사용자 나이는 22세임

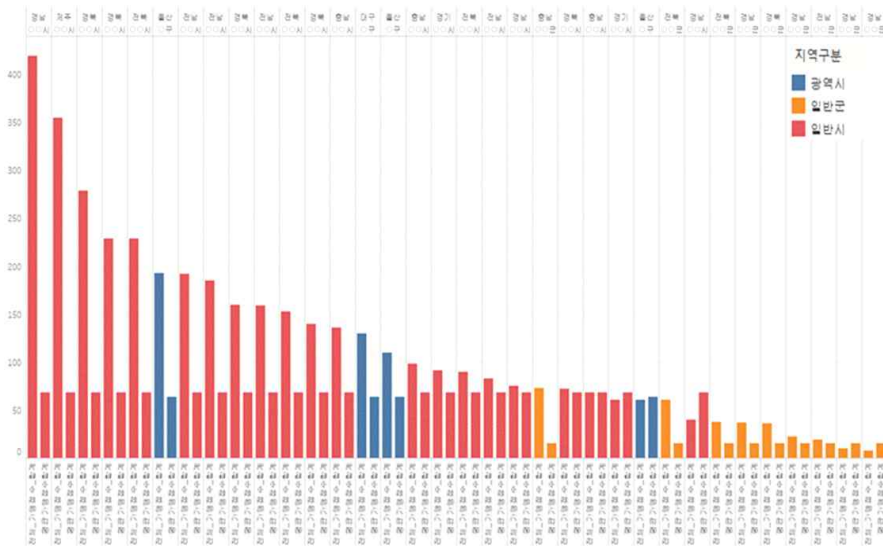
<표 3-4> 발급 카드 수 대비 미 사용 카드 수가 적은 지역 비교(높은 사용)

연도	2016	2017
시군구	전북 ○○시	울산 ○구
주민센터	○○동주민센터	○○ 2 동주민센터
미사용카드수 /카드발급수	0.4%	0.5%
카드사용 평균나이 (세 )	51.3	41.2
사용건수 _합계	837	762
사용금액 _Sum(원 )	13,262,640	12,824,830
보조금 사용액 _Sum(원 )	13,236,630	12,757,580
개인충전금 사용액 _Sum(원 )	26,010	67,250
보조금 잔액 _Min_Sum (원 )	678,710	516,370
개인충전금 잔액 _Min_Sum (원 )	35,543	9,000
카드별 사용건수 _Sum(건 )	843	762
사용건수 /카드발급수 _Mean(건 )	2.8	4.3
최종잔액 _Sum (원 )	50,000	60,000
지역구분	일반시	광역시
미사용자 _나이 _Mean(세 )	84	22
미사용카드수 _Sum	1	1
사용카드수 _합	259	200
총 사용금액 /카드발급수 (원 /수 )	51,010	63,805
카드발급수	260	201

- 표 3-4을 통해 두 주민센터의 특징을 살펴보면 총 사용금액은 약 1,326만 원 정도이고, 카드를 사용한 평균 나이는 40~50대이며, 미사용자의 나이는 84세 고령자와 22세 청년임
- 미사용자의 나이로 볼 때 카드 발급 후 카드사용에 대해서 인지를 못 했을 가능성이 높으므로 향후 홍보 및 카드사용 문자 등에 대한 알림서비스를 통해 지속적 관리가 필요함
- 사용 카드 수가 많은 지역과 적은 지역이 해당 지역 사용분야(업종) 수와 관계가 있는지 살펴보기 위해서 미사용 카드 수와 사용분야(업종) 수의 비교 분석을 진행함

- 각 지역 내 전체 사용분야(업종) 수 합계와 지역구분별(특별시, 광역시, 일반시, 일반군) 평균 사용분야(업종) 수 합계와 관계를 살펴보기 위해 그래프로 시각화함
- 그 결과, 카드 사용이 많은 지역으로 도출된 37개 시군구들 중 30개(81%)<sup>4)</sup>는 평균 사용분야(업종) 수 보다 많은 사용분야(업종) 들이 분포되어 있었음
- 이는 “카드 사용이 많은 지역은 사용분야(업종) 수가 많다”라는 것을 알 수 있음
- 다음으로 발급 카드 수 대비 미사용 카드 수가 많은 '15년도, '16년도, '17년도 상위 20곳을 도출하면 표 3-5, 3-6, 3-7과 같음

[그림 3-24] 지역내 전체 사용분야(업종) 수와 지역구분별 평균 사용분야(업종) 비교분석  
(카드 사용률이 높은 지역)



4) 전체 사용분야(업종) 수는 각 지역별 사용분야(업종) 수의 총합을 뜻하며, 평균 사용분야(업종) 수는 지역구분별(광역시, 일반군, 일반시, 특별시)의 평균 사용분야(업종) 수의 합계를 뜻함

&lt;표 3-5&gt; '15년 기준 사용카드 수 가 적은 주민센터 (상위 20위)

시군구	주민센터	미사용카드수 / 카드발급수
전남 ○○시	○○동 주민센터	26.4%
전남 ○○시	○○ 주민자치센터	26.5%
경남 ○○군	○○면사무소	26.7%
서울 ○구	○○동주민센터	27.3%
전남 ○○군	○○면사무소	27.4%
인천 ○○군	○○면사무소	28.8%
전남 ○○시	○○면 주민센터	29.2%
강원 ○○군	○면사무소	29.5%
제주 ○○시	○○면사무소	30.0%
경남 ○○군	○○면사무소	30.2%
전남 ○○군	○○면사무소	31.0%
서울 ○구	○○동주민센터	31.1%
강원 ○○시	○○면사무소	31.3%
충남 ○○군	○○면사무소	33.7%
전남 ○○시	○○면 주민센터	34.2%
전남 ○○군	○○면사무소	34.8%
전남 ○○군	○○면사무소	35.3%
강원 ○○시	○○면사무소	38.9%
전남 ○○군	○○면사무소	39.6%
강원 ○○시	○○면사무소	40.0%

&lt;표 3-6&gt; '16년도 기준 사용카드 수 가 적은 주민센터 (상위 20위)

시군구	주민센터	미사용카드수 / 카드발급수
충북 ○○군	○○면사무소	22.8%
전남 ○○군	○○면사무소	22.9%
강원 ○○군	○○○면사무소	23.0%
전남 ○○군	○○면사무소	23.1%
강원 ○○군	○○읍 행정복지센터	23.8%
충남 ○○시	○○면주민센터	24.1%
전남 ○○군	○○면사무소	24.3%
충남 ○○시	○○면사무소	25.0%
전남 ○○군	○○면사무소	25.3%
충남 ○○군	○○면 행정복지센터	25.4%
경기 ○○시	○○면사무소	26.5%
전남 ○○군	○○면사무소	27.8%
경북 ○○시	○○면사무소	28.6%
전남 ○○군	○○면사무소	30.1%
경북 ○○시	○○면주민센터	30.4%
인천 ○○군	○○면주민자치센터	33.3%
충북 ○○군	○○면사무소	35.3%
전남 ○○군	○○면사무소	35.5%
충북 ○○군	○○면사무소	35.7%
인천 ○○군	○○면사무소	42.1%

&lt;표 3-7&gt; '17년도 기준 사용카드 수 가 적은 주민센터 (상위 20위)

시군구	주민센터	미사용카드수 / 카드발급수
경북 ○○시	○○면주민센터	33.7%
강원 ○○군	○○○면사무소	34.0%
경북 ○○군	○○면주민센터	34.2%
강원 ○○시	○○면사무소	35.4%
전남 ○○군	○○면사무소	35.6%
충북 ○○시	○○면주민센터	35.9%
전남 ○○군	○○면사무소	36.3%
전남 ○○군	○○면사무소	36.9%
충북 ○○시	○○ 2 동주민센터	37.5%
충북 ○○시	○○면사무소	37.8%
전남 ○○군	○○면사무소	37.9%
경북 ○○군	○○면주민센터	38.4%
충북 ○○시	○○면주민센터	39.0%
충북 ○○시	○○면사무소	39.4%
충북 ○○시	○○면주민센터	40.8%
충북 ○○군	○○면사무소	42.4%
강원 ○○시	○○면사무소	44.9%
충북 ○○군	○○읍사무소	44.9%
경북 ○○군	○○면주민센터	47.7%
전남 ○○군	○○면사무소	48.0%

- '16년도 카드사용이 미비한 전남 ○○시의 ○○면사무소와 '17년도 미비한 강원 ○○군의 ○○○면사무소, 강원 ○○시의 ○○면사무소의 발급 카드 수 대비 미사용 카드 수를 비교하였음
- 전남 ○○시의 ○○면사무소의 경우 카드사용 평균 나이가 52.6세, 미사용나이 67.2세, 미사용 카드 수 100개, 발급카드 수 182개, 총 사용금액/카드발급수(원/수) 29,565원으로 도출됨. 182개 발급카드 중 100개가 미사용 되어 낮은 카드 사용률을 보이고 있음
- 강원 ○○군의 ○○○면사무소의 경우 카드사용 평균 나이가 37.2세, 미사용 평균 나이 51.9세, 미사용 카드 수 17개, 발급카드 수 33개, 총 사용금액/카드발급수(원/수) 46,818원으로 도출됨. 33개 발급카드중 17개가 미사용 되어 낮은 카드 사용률을 보이고 있음

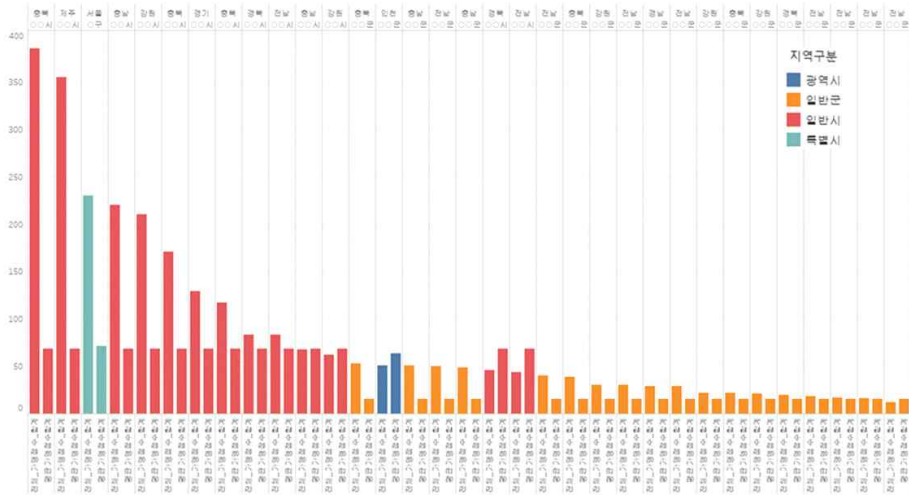
- 강원 ○○시의 ○○면사무소의 경우 카드사용 평균 나이가 63.6세, 미사용 나이 70.2세, 미사용 카드 수 22개, 발급카드 수 27개, 총 사용금액/카드발급수(원/수) 33,763원으로 도출됨. 27개 발급카드 중 22개가 미사용 되어 낮은 카드 사용률을 보이고 있음
- 이를 통해 미사용 평균 연령대가 모두 고령자임을 알 수 있고 사용금액도 타지역 평균 이하로 나타나 이들 지역에 대한 홍보를 강화하거나 고령자 친화적인 사용분야(업종)를 신설하고 이를 사용하게 유도하는 정책이 필요

&lt;표 3-8&gt; 발급 카드 수 대비 미사용 카드 수 비교 (낮은 사용)

연도	2016	2017	2017
시군구	전남 ○○군	강원 ○○군	강원 ○○시
주민센터	○○면사무소	○○○면사무소	○○면사무소
미사용카드수 /카드발급수	35.5%	34.0%	44.9%
카드사용 평균나이 (세 )	52.6	37.2	63.6
사용건수 _합계	510	167	44
사용금액 _Sum(원 )	8,337,330	2,340,912	1,654,400
보조금 사용액 _Sum(원 )	8,325,930	2,340,912	1,654,400
개인충전금 사용액 _Sum(원 )	11,400	0	0
보조금 잔액 _Min_Sum ( 원 )	1,078,570	284,828	85,600
개인충전금 잔액 _Min_Sum ( 원 )	0	0	0
카드별 사용건수 _Sum( 건 )	510	167	44
사용건수 /발급카드수 _Mean(건 )	3.8	6.2	1.4
최종잔액 _Sum ( 원 )	5,050,000	1,020,000	1,320,000
지역구분	일반군	일반군	일반시
미사용자 _나이 _Mean(세 )	67.2	51.9	70.2
미사용카드수 _Sum	100	17	22
사용카드수 _합	182	33	27
총 사용금액 /카드발급수 (원 /수 )	29,565	46,818	33,763
카드발급수	282	50	49

- 즉, 표 3-8을 통해 세 곳의 주민센터의 특징을 살펴보면 사용금액은 200만 원, 800만 원대였고, 카드를 사용한 평균나이는 30대 후반 ~ 60대 초반이며, 미사용자의 나이는 대부분 60대 후반 이후로 고령자인 것을 알 수 있음
- 카드 사용이 낮은 지역으로 도출된 34개 시군구들 중 27개(79%)는 평균 사용분야(업종) 수 보다 많은 사용분야(업종) 들이 분포되어 있지만 일반군의 경우 53%가 평균 사용분야(업종) 수 보다 적은 사용분야(업종)들이 분포되어 있음
- 이는 “카드 사용이 적은 지역은 사용분야(업종) 수 가 적다”라는 것을 알 수 있음

[그림 3-25] 지역 내 전체 사용분야(업종) 수와 지역구분별 평균 사용분야(업종) 비교분석  
(카드 사용률이 낮은 지역)



- 다음으로, 2015~2017년 광역시, 일반군, 일반시, 특별시의 발급 카드 대비 카드사용 건 수 비교를 진행함으로써 카드사용의 활성화를 살펴보고자 함
- 광역시의 경우 '16년 2.6건에서 '17년 2.7건으로 4% 성장을, 일반군의 경우 '16년 2.5건에서 '17년 2.7건으로 9% 성장을, 일반시의 경우 '16년 2.6건에서 '17년 2.7건으로 5% 성장을, 특별시의 경우 '16년 2.6건에서 '17년 2.7건으로 1%로 성장하여 전체적으로 사용률이 높아짐을 보이고 있음

[그림 3-26] 발급 카드 대비 사용건 수 비교

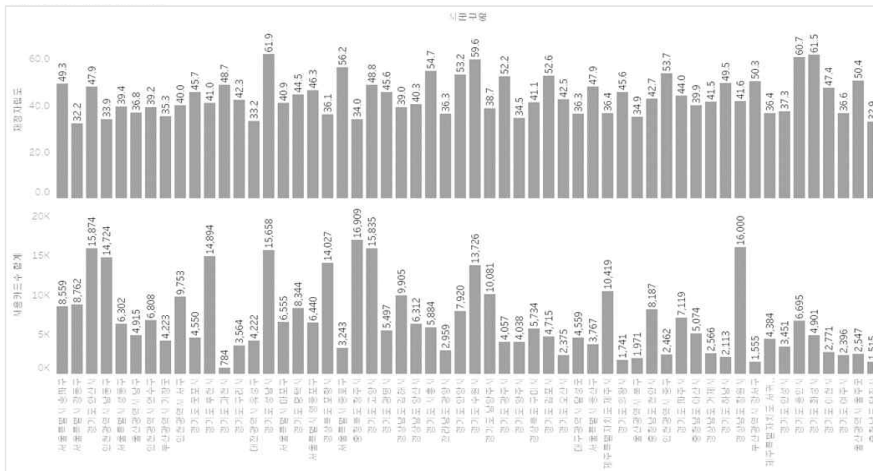


- 전체 243개 지자체 중 17개 광역 시도 및 데이터 오류가 있는 지자체를 제외한 222개를 대상으로 지자체 재정자립도<sup>5)</sup>를 살펴보았을 때 상관관계는 0.171로 나타났다
- 그림 3-27과 3-28을 통해 지자체 재정자립도 4분위 수 중 상위 50%를 살펴보면 재정자립도와 카드발급 수의 합계에 대한 상관관계는 0.171로 연관성이 낮은 것으로 도출됨

[그림 3-27] 상위 50% 재정자립도 상관관계

		재정자립도	카드발급수
재정자립도	Pearson 상관	1	.171*
	유의확률(양측)		.012
	N	214	214
카드발급수	Pearson 상관	.171*	1
	유의확률(양측)	.012	
	N	214	214

[그림 3-28] 상위 50% 재정자립도



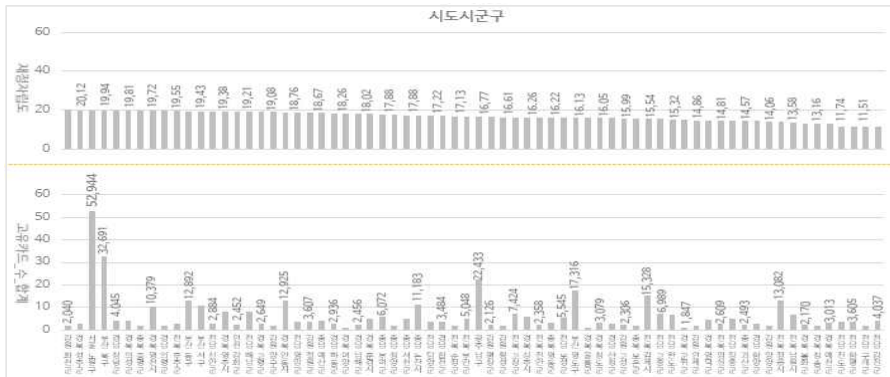
5) 재정자립도 : 지방정부가 재정 활동에 필요한 자금을 어느 정도나 자체적으로 조달하고 있는가를 나타내는 지표

- 그림 3-29와 3-30을 통해 지자체 재정자립도 4분위 수 중 중위 50% 이하를 살펴보면 재정자립도와 카드발급 수 합계의 Pearson상관계수를 살펴보면 0.448로 상관관계가 높은 것을 알 수 있음

[그림 3-29] 중위 50% 이하 재정자립도 상관관계

		재정자립도	카드발급수
재정자립도	Pearson 상관	1	.448
	유의확률(양측)		.000
	N	111	111
카드발급수	Pearson 상관	.448	1
	유의확률(양측)	.000	
	N	111	111

[그림 3-30] 중위 50% 이하 재정자립도



- 중위 50%(하위 111개 지자체)는 지자체 재정자립도가 낮은 지역일수록 발급 수가 낮은 특성을 보임
- 카드사용 활성화를 위해서는 통합문화이용권(문화누리카드) 수혜자(장애인·고령자 등)정보를 통한 맞춤형 지원과 지방비 매칭을 통해 콘텐츠 및 홍보의 다양화를 제공하는 등 예산 정책에 대한 고려가 필요함을 의미함
- 현재 시도별 재정자립도를 고려하여 지방비 매칭비율<sup>6)</sup>을 적용하고 있으나 향후 시군구별 재정자립도까지 감안한 매칭비율을 적용하는 것이 효율적인 것인 것으로 판단함

6) 7:3 범위내에서 광역시도별로 재정자립도에 따라 매칭 비율을 다르게 하고 있음

### － 문화누리카드 문의사항에 따른 단어빈도 분석

- 문화누리카드 문의사항 테이블은 총 14,329건의 컬럼으로 이루어져 있으며, 총 수집 시기는 2016년 11월 28일부터 2017년 12월 12일까지임
- 구분 컬럼이 문의사항으로 분류되어있는 9,689건을 대상으로 11가지 유형으로 분류 후 질문 유형별 특성 분석을 진행하였음

<표 3-9> 문화누리카드 문의사항 테이블 데이터 구분 사항

NO.	구분	빈도 (건 )
1	답변	4,542
2	메모	98
3	문의사항	9,689
총계		14,329

- Python을 활용한 단어 빈도 분석과 Konlpy라이브러리 사용하여 명사를 수집한 후, 단어의 빈도수를 계산하여 정보를 기록함. 아래 표 3-10에 있는 소스는 실제로 단어 빈도 분석을 수행한 소스임

<표 3-10> 단어 빈도 분석 소스

```
from konlpy.tag import Twitter
from collections import Counter
def get_tags(text, ntags=50):
    splitter = VoC()
    nouns = splitter.nouns(text)
    count = Counter(nouns)
    return_list = []
    for n, c in count.most_common(ntags):
        temp = {'tag': n, 'count': c}
        return_list.append(temp)
    return return_list
def main():
    text_file_name = "문화누리카드 문의사항.txt"
    noun_count = 200
    output_file_name = "wellchon_portal_netizen_count.txt"
    open_text_file = open(text_file_name, 'r', -1, "utf-8")
    text = open_text_file.read()
    tags = get_tags(text, noun_count)
    open_text_file.close()
    open_output_file = open(output_file_name, 'w', -1, "utf-8")
    for tag in tags:
```

```

noun = tag['tag']
count = tag['count']
open_output_file.write('{} {}{}\n'.format(noun, count))
open_output_file.close()
if __name__ == '__main__':
    main()

```

- Excel의 필터 및 부분합을 통해 11가지 문의사항 질문 유형을 분석함. 11가지 질문 유형은 사용분야(업종) 위치, 사용분야(업종) 사용, 기타, 발급자격 및 서류, 비밀번호 변경, 삭제, 시스템 오류, 자부담 충전, 잔액 조회, 카드 분식 및 등록, 카드 합산으로 총 9,689건에 대한 분석 결과를 도출하였음 (표 3-11 참고)

<표 3-11> 문화누리카드 문의사항 질문 유형

NO.	구분	빈도 (건)
1	삭제	5,381
2	기타	2,252
3	사용분야(업종) 사용	609
4	비밀번호 변경	597
5	카드 분실 및 등록	198
6	자부담 충전	194
7	발급자격 및 서류	167
8	잔액 조회	130
9	사용분야(업종) 위치	50
10	시스템오류	47
11	카드합산	38
총계		9,689

- 질문 유형 중 ‘삭제’로 분류되는 질문들은 문화누리카드의 문의사항과는 관련이 없는 것으로, 본 분석에서는 제외하였음. 단어 빈도 분석은 문의사항 전체에 대한 단어 빈도 분석을 실시하였으며, 키워드 빈도수를 통해 문의사항의 추세를 파악하였음

&lt;표 3-12&gt; 문화누리카드 문의사항 단어 빈도 분석 결과

NO.	키워드	빈도 (건 )	NO.	키워드	빈도 (건 )
1	카드	2,159	11	문자	227
2	연락	2,099	12	영화	210
3	문화	1,267	13	변경	206
4	누리	1,188	14	등록	185
5	비밀번호	768	15	번호	179
6	사용	588	16	발급	161
7	제목	460	17	방법	160
8	충전	460	18	오늘	160
9	결제	281	19	온라인	139
10	잔액	245	20	가맹	134

- 문화누리카드 문의사항' 전체를 대상으로 빈도 분석을 진행한 결과인 표 3-12를 참고하면 '카드' 2,159건, '연락' 2,099건, '문화' 1,267건, '누리' 1,188건, '비밀번호' 768건 순으로 나타났으며, '온라인' 139건, '가맹' 134건으로 낮은 빈도를 보였음
- '사용' 588건, '충전' 460건, '결제' 281건, '잔액' 245건 등의 키워드로 문화누리 카드 결제 방안에 관한 문의가 많았음을 알 수 있으며, '등록' 185건, '발급' 161건, '온라인' 139건 등의 키워드로 카드 발급 및 수령 후 등록 과정에 관한 문의가 많았고, '영화' 210건 등의 키워드로 문화누리 카드의 사용처에 관한 문의가 많음을 알 수 있음
- 질문유형이 '사용분야(업종) 위치'인 문의사항 중 명사를 추출하여 분석을 진행하였으며, 본 주제와 관련 없는 내용은 제외하였음

&lt;표 3-13&gt; 사용분야(업종) 위치 단어 빈도 분석 결과

NO.	키워드	빈도 (건 )	NO.	키워드	빈도 (건 )
1	카드	17	11	숙박	2
2	문화	14	12	여수	2
3	누리	13	13	도서	1
4	위치	6	14	정보	1

5	서점	5	15	운동	1
6	용품	3	16	티켓	1
7	지역	3	17	테마파크	1
8	오프라인	3	18	하남스타필드	1
9	수영장	2	19	오션월드	1
10	체육	2	20	케이블카	1

- 문화누리카드 ‘사용분야(업종) 위치’를 대상으로 빈도 분석을 진행한 결과 ‘카드’ 17건, ‘문화’ 14건, ‘위치’ 6건, ‘서점’ 5건, ‘용품’ 3건, ‘지역’ 3건, ‘오프라인’ 3건 순으로 나타났으며, ‘수영장’ 2건, ‘체육’ 2건의 빈도수를 보였음 (표 3-13 참고)
- ‘사용분야(업종) 위치’ 대상에서는 서점, 체육시설, 오프라인 사용분야(업종) 등의 위치를 묻는 질문이 많은 것으로 나타났으며, 이는 내 주변에 위치해 있는 사용분야(업종) 찾기 등 주변 사용분야(업종) 현황에 대한 정보가 필요한 것으로 나타났음
- 질문유형이 ‘사용분야(업종) 사용’인 문의사항 중 명사를 추출하여 분석을 진행하였으며, 본 주제와 관련 없는 내용은 제외하였음

<표 3-14> 사용분야(업종) 사용 단어 빈도 분석 결과

NO.	키워드	빈도 ( 건 )	NO.	키워드	빈도 ( 건 )
1	사용	158	11	사용	26
2	결제	90	12	등록	24
3	영화	68	13	롯데	23
4	가맹	66	14	서점	22
5	할인	47	15	구매	21
6	책	42	16	오프라인	21
7	구입	40	17	교보	20
8	온라인	38	18	고속버스	19
9	예매	33	19	결제	19
10	인터넷	31	20	홈페이지	17

- 사용분야(업종) 사용 단어 빈도 분석 결과 ‘사용’ 158건, ‘결제’ 90건, ‘영화’ 68건, ‘가맹’ 66건, ‘할인’ 47건순으로 높게 나타났으며, ‘구매오프라인’ 21건, ‘교보’ 20건, ‘고속버스결제’ 19건, ‘홈페이지’ 17건으로 낮은 순으로 나타났음 (표 3-14 참고)
- ‘할인’ 47건, ‘구입’ 40건, ‘온라인’ 38건, ‘예매’ 33건, ‘인터넷’ 31건, ‘사용’ 26건, ‘등록’ 24건 등의 키워드로 보았을 때 문화누리 카드의 사용 방안에 대한 문의가 많았음
- 수요에 대한 문의는 ‘영화’ 68건으로 가장 많았고, 책 42건, ‘서점’ 22건 등의 키워드로써 도서에 관한 문의가 두 번째로 많았음
- ‘고속버스’ 19건, 표 3-14에 제시된 20개 단어이외에 ‘버스’, ‘자전거’, ‘기차’ 등의 키워드가 나타났으며, 이는 이동수단에 관한 문의가 세 번째로 많음을 알 수 있음
- 질문 유형이 ‘기타’인 문의사항 중 명사를 추출하여 분석을 진행하였으며, 본 주제와 관련 없는 내용은 제외하였음

&lt;표 3-15&gt; 기타 단어 빈도 분석 결과

NO.	키워드	빈도 ( 건 )	NO.	키워드	빈도 ( 건 )
1	연락	1,283	11	돈	41
2	충전	163	12	기간	41
3	문자	88	13	가요	40
4	결제	82	14	영화	38
5	번호	65	15	온라인	36
6	잔액	61	16	홈페이지	35
7	등록	55	17	주민	34
8	발급	48	18	전화	34
9	비밀번호	44	19	금액	34
10	신청	41	20	인터넷	31

- 기타 문의사항 단어 빈도 분석 결과 연락이 1,283건으로 가장 많았고, 충전 163건, 문자 88건, 결제 82건, 번호 65건으로 각각 높은 키워드 빈도를 보이는 것으로 나타났음. 온라인 36건, 홈페이지 35건, 주민-전화금액 34건, 인터넷 31건으로 각각 낮은 키워드 빈도를 보였음 (표 3-15 참고)

- 문자 연락을 원하는 사람들의 문의가 많은 것으로 보아 전화로(스마트폰) 문의사항을 전달하고자 하는 수혜자가 많음을 알 수 있었음(그림 3-31 참고)

[그림 3-31] 전화문의사항(일부분)

1	2	3	A	B	C	D	E
			번호	질문유형	구분	성명내용 및 답변	시간
	667	59503	기타	문의사항	연락 바랍니다.		2017-12-11 08:49:41
	674	59442	기타	문의사항	연락 바랍니다.		2017-12-09 10:45:27
	675	59435	기타	문의사항	연락 바랍니다.		2017-12-09 08:30:55
	678	59389	기타	문의사항	연락 바랍니다.		2017-12-06 18:55:01
	679	59379	기타	문의사항	연락 바랍니다.		2017-12-06 16:00:33

- 자부담 충전금 및 결제, 잔액 등 문의 사항이 많았고, 이는 온라인에서 제공되고 있는 서비스에 대해 수혜자 인식이 높지 않음을 알 수 있었으며, 등록 및 발급에 관련된 문의사항이 많은 것은 문화누리 카드 발급 과정 및 혜택에 대한 인식이 아직 미비함을 알 수 있음
- 질문 유형이 ‘발급자격 및 서류 문의사항’인 문의사항 중 명사를 추출하여 분석을 진행하였으며, 본 주제와 관련 없는 내용은 제외하였음

&lt;표 3-16&gt; 발급자격 및 서류 문의사항 단어 빈도 분석 결과

NO.	키워드	빈도 (건 )	NO.	키워드	빈도 (건 )
1	발급	53	16	센터	9
2	충전	51	17	등록	7
3	재	21	18	수급	7
4	사용	17	19	본인	6
5	신청	17	20	돈	6
6	올해	16	21	수령	6
7	대상자	15	22	차상위	6
8	주민	13	23	한부모	5
9	기간	12	24	온라인	5
10	재발	12	25	기초	4
11	대상	12	26	금액	4
12	가요	10	27	지원	4
13	인터넷	10	28	방법	4
14	제목	9	29	자격	4
15	교육	9	30	잔액	3

- ‘발급자격 및 서류’문의사항 단어 빈도 분석 결과 ‘발급’ 53건, ‘충전’ 51건, ‘재가’ 21건, ‘사용’ 17건, ‘신청’ 17건으로 각각 높은 빈도를 보였으며, 기초금액 지원방법자각잔액은 각각 4건으로 낮은 빈도를 보였음 (표 3-16 참고)
- ‘발급’ 53건, ‘신청’ 17건, ‘대상자’ 15건, ‘자격’ 4건, ‘방법’ 4건 등의 키워드 순으로 발급 대상인지를 물어보는 문의가 많음을 알 수 있으며, ‘센터’ 9건, ‘본인’ 6건, ‘수령’ 6건 등의 키워드에서 제시되었듯이 카드 발급자들의 수령 방법에 관한 공지가 지속적으로 필요할 것으로 판단됨
- 질문 유형이 ‘비밀번호 변경’인 문의사항 중 명사를 추출하여 분석을 진행하였으며, 본 주제와 관련 없는 내용은 제외하였음

&lt;표 3-17&gt; 비밀번호 변경 단어 빈도 분석 결과

NO.	키워드	빈도 ( 건 )	NO.	키워드	빈도 ( 건 )
1	비밀번호	599	16	온라인	23
2	변경	134	17	영화	21
3	오류	71	18	예매	17
4	정지	57	19	발급	16
5	비번	55	20	인터넷	15
6	사용	52	21	확인	15
7	초과	52	22	분실	14
8	방법	47	23	회원	14
9	결제	43	24	구매	13
10	오류로	40	25	주민	8
11	횟수	39	26	인증	7
12	제목	36	27	책	7
13	번호	28	28	찾기	7
14	홈페이지	27	29	로그인	7
15	설정	26	30	모바일	6

- ‘비밀번호’가 599건으로 가장 많았고, ‘변경’ 134건, ‘오류’ 71건, ‘정지’ 57건, ‘비번’ 55건 순으로 단어 빈도가 높게 나타났으며, ‘주민’ 8건, 인증책찾가.로그인이 각각 7건, ‘모바일’ 6건으로 단어 빈도가 낮게 나타남(표 3-17 참고)
- 3회 이상 비밀번호가 틀릴 시 카드가 정지되고, 비밀번호를 변경해야 사용할 수 있는 특성과 관련된 ‘정지’, ‘비번’, ‘초과’, ‘오류’, ‘분실’ 등의 키워드를 확인할 수 있고 ‘정지’로 인한 비밀번호 변경에 관한 문의가 많음을 알 수 있었음
- 질문 유형이 ‘시스템 오류’, ‘잔액조회’인 문의사항 중 명사를 추출하여 분석을 진행하였으며, 본 주제와 관련 없는 내용은 제외하였음

&lt;표 3-18&gt; 시스템 오류, 잔액조회 단어 빈도 분석 결과

시스템 오류 문의사항 단어 빈도			잔액조회 문의사항 단어 빈도		
NO.	키워드	빈도 ( 건 )	NO.	키워드	빈도 ( 건 )
1	충전	9	1	잔액	93
2	결제	9	2	카드	46
3	홈페이지	8	3	조회	30
4	비밀번호	8	4	확인	27
5	오류	8	5	얼마	26
6	인증	6	6	금액	11
7	서비스	4	7	문자	8
8	등록	4	8	사용	8
9	번호	4	9	결제	5
10	확인	4	10	얼마나	5
11	정지	4	11	제목	4
12	사이트	3	12	문의	4
13	잔액	3	13	방법	4
14	로그인	2	14	전화	3
15	비정상	2	15	충전	3

- 시스템 오류와 관련된 문의사항 단어 빈도 분석 결과 충전-결제는 각 9건, 홈페이지비밀 번호-오류는 각 8건으로 높게 도출되었고, 로그인-비정상은 각 2건으로 낮게 도출되었음 (표 3-18 참고)
- ‘비밀번호’, ‘오류’, ‘인증’, ‘로그인’, ‘비정상’등의 시스템 인증과 관련된 오류가 많음을 알 수 있음
- ‘잔액조회’와 관련된 문의사항 단어 빈도 분석 결과 ‘잔액’ 93건, ‘카드’ 46건, ‘조회’ 30건, ‘확인’ 27건, ‘얼마’ 26건이 높게 도출되었고, 전화-충전은 각 3건으로 낮게 도출되었음
- 카드에 잔액이 얼마나 남았는지에 관한 문의와 남은 잔액의 사용 방법에 관한 문의가 많음을 알 수 있음
- 질문 유형이 ‘자부담 충전’인 문의사항 중 명사를 추출하여 분석을 진행하였으며, 본 주제와 관련 없는 내용은 제외하였음

<표 3-19> 자부담 충전 단어 빈도 분석 결과

NO.	키워드	빈도 ( 건 )	NO.	키워드	빈도 ( 건 )
1	충전	158	16	부담	11
2	계좌	44	17	시간	8
3	재	42	18	언제	8
4	입금	42	19	농협	8
5	가상	35	20	영화	8
6	금액	21	21	환불	7
7	잔액	21	22	번호	7
8	개인	20	23	본인	7
9	돈	18	24	홈페이지	6
10	방법	14	25	할인	6
11	한도	14	26	발급	5
12	인터넷	14	27	동사무소	5
13	사용	13	28	온라인	5
14	제목	13	29	구입	4
15	가요	13	30	현질	2

- ‘자부담 충전’ 문의사항 단어 빈도 분석 결과 ‘충전’ 158건으로 가장 높았으며, ‘계좌’ 44건, ‘입금’ 42건, ‘가상’ 35건 등의 순으로 높게 나타났음 (표 3-19 참고)
- 그와 반대로 ‘할인’ 5건, ‘발급동사무소온라인’ 5건, ‘구입’ 4건 순으로 낮게 도출되었음
- 자부담 충전 방안 및 그와 관련된 가상계좌에 관한 문의가 많은 것은 관련 정보에 대한 공지가 수혜자에게 지속적으로 필요한 것으로 판단됨
- 또한, 카드 사용 시 한도 초과에 관한 문의와 카드 환불 일정에 관한 문의가 많은 것으로 보아 카드한도 정보를 알 수 있는 시스템이 필요함
- 질문 유형이 ‘카드 분실 및 등록’인 문의사항 중 명사를 추출하여 분석을 진행 하였으며, 본 주제와 관련 없는 내용은 제외하였음

<표 3-20> 카드 분실 및 등록 단어 빈도 분석 결과

NO.	키워드	빈도 (건 )	NO.	키워드	빈도 (건 )
1	카드	181	16	정지	7
2	분실	76	17	충전	6
3	등록	63	18	확인	5
4	사용	42	19	신청	5
5	재발	30	20	교부	5
6	신고	18	21	주민	5
7	수령	17	22	전화	5
8	발급	15	23	동사무소	5
9	결제	10	24	센터	4
10	잔액	9	25	조회	4
11	인터넷	9	26	비밀번호	4
12	농협	8	27	상담	4
13	온라인	8	28	승인	3
14	홈페이지	7	29	회원	3
15	가요	7	30	문자	3

- 카드 분실 및 등록 문의사항 단어 빈도 분석 결과 ‘카드’ 181건으로 가장 높게 나타났으며, ‘분실’ 76건, ‘등록’ 63건, ‘사용’ 42건, ‘재발’ 30건 등의 순으로 높은 빈도를 보였으며, ‘승인’ 과 ‘회원’, ‘문자’는 각 3건으로 낮은 빈도를 보였음 (표 3-20 참고)
- ‘분실’, ‘신고’ 등의 키워드를 통해 카드 분실 후 분실 신고에 관한 문의사항이 많음을 알 수 있으며, ‘등록’, ‘사용’, ‘수령’, ‘발급’등의 키워드를 통해 분실 신고 후 재발급 과정에 관한 문의 사항이 많음을 알 수 있음
- 질문 유형이 ‘카드 합산’인 문의사항 중 명사를 추출하여 분석을 진행 하였으며, 본 주제와 관련 없는 내용은 제외하였음

&lt;표 3-21&gt; 카드 합산 단어 빈도 분석 결과

NO.	키워드	빈도 (건 )	NO.	키워드	빈도 (건 )
1	카드	42	16	인증서	2
2	합산	39	17	인터넷	2
3	방법	9	18	인증	2
4	가족	8	19	구성원	2
5	사용	6	20	카트	1
6	잔액	6	21	비밀번호	1
7	세대	3	22	요금	1
8	문의	3	23	아빠	1
9	충전	3	24	아버지	1
10	제목	3	25	언니	1
11	자녀	2	26	미성년	1
12	안내	2	27	개인	1
13	온라인	2	28	동생	1
14	동사무소	2	29	엄마	1
15	공인	2	30	신청	1

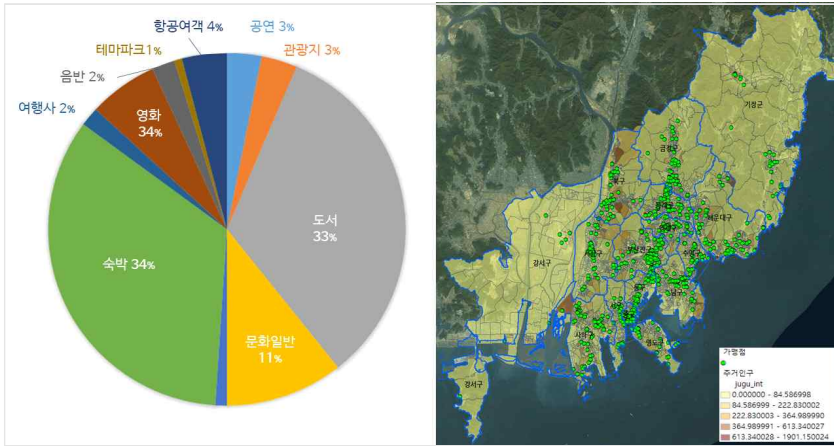
- 카드 합산 문의사항 단어 빈도 분석 결과 ‘카드’ 42건, ‘합산’ 39건으로 다른 단어에 비해 매우 높은 빈도를 나타냈고, 그다음으로 ‘방법’ 9건, ‘가족’ 8건, ‘사용’과 ‘잔액’이 각 6건으로 높게 도출되었음 (표 3-21 참고)
- 카드 합산 방법에 관한 문의가 많은 것을 확인할 수 있으며, 이에 따른 카드 합산 방안에 대한 방법을 재 공지 또는 활성화 시킬 필요가 있음

- 가족, 자녀, 구성원, 아빠, 아버지, 언니 등의 가족과 관련된 키워드를 통하여 가족에 대한 카드 합산 방안의 문의가 많음을 알 수 있으며, 구성원의 합산 문의사항을 줄이기 위하여 개인별 카드 발급 이후 합산이 아닌 구성원 단위의 카드 발급이 필요한 것으로 도출되었음

#### □ 사용분야(업종) 공간 분석

- 사용분야(업종) 분석에서 동일 사용분야(업종) 밀집도를 분석하기 위해 커널분석을 활용하였음
- 커널분석은 커널함수를 사용하여 공간상의 점(Point) 분포패턴을 효과적으로 나타낼 수 있는 방법으로, 커널함수는 일정한 분포 반경(Bandwidth)안에 포함하는 점 데이터 밀도를 측정하여 커널함수 K로 표현함
- 커널밀도 분석은 점 객체가 위치한 지점에서 가장 높은 값이 부여되고 이후 거리가 멀어지게 됨에 따라 그 값이 작아져 설정한 반경의 경계를 벗어나면 그 값이 부여되지 않는 특성이 있음
- 부산시를 대상으로 공간정보와 GIS기술을 활용하여 사용분야(업종)간의 분포와 확대 대상지를 도출하는 공간 빅데이터 분석 방법론과 커널분석을 적용하였음
- 먼저, 부산시 내 전체 사용분야(업종)은 공연 27개, 관광지 28개, 도서 284개, 문화일반 92개, 문화체험 9개, 숙박 294개, 여행사 15개, 영화 54개, 음반 18개, 테마파크/레저 6개, 항공여객고속버스 렌터카가 35개로 총 862개의 사용분야(업종)이 있으며(그림 3-32 참고) 전국사업체조사 위치, 거주인구, 유동인구, 카드매출, 사용분야(업종) 위치, 문화누리 카드 사용카드 수를 공간 Join하여 진행하였음

[그림 3-32] (좌) 부산 전체 사용분야(업종) 분야별 비율 (우) 주거인구와 사용분야(업종) 분포도



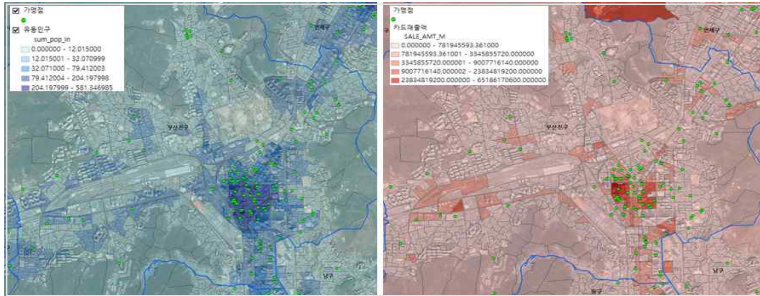
- 커널밀도 분석을 활용해 동일 사용분야(업종)들이 지역 내 분포와 이에 따른 영향력 분포를 분석함으로써 추가적인 사용분야(업종) 확대 계획을 수립하고, 숙박 사용분야(업종)는 부산 내 상대적으로 분포가 다양하지만, 도서 및 문화일반 사용분야(업종)는 부산 진구 지역에 밀집하여 있기 때문에 타 상권 방안을 확산할 필요가 있음 (그림 3-33 참고)

[그림 3-33] 동일 사용분야(업종)의 밀집도 분석



- 그림 3-34와 같이 외부 데이터와 연계하여 유동인구 데이터와 카드매출 데이터를 기반으로 사용분야(업종)의 위치를 분석하였음

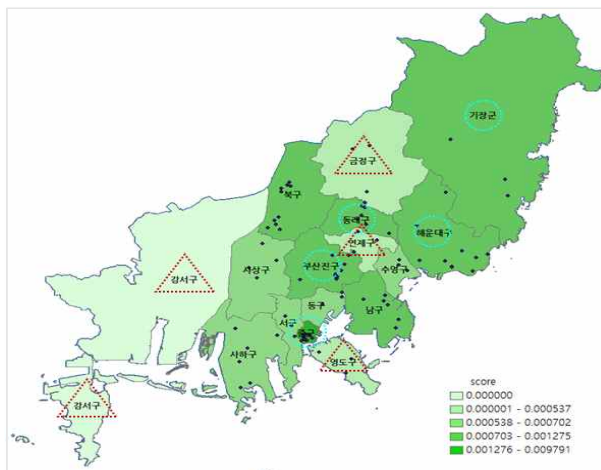
[그림 3-34] 부산 진구 일대 사용분야(업종) 분포



- 대체로 사용분야(업종)는 유동인구가 많을수록 사용분야(업종)의 분포 비율이 높고, 사업자의 카드 매출액 기반 사용분야(업종)의 위치를 분석하면 카드 매출이 높은 지역일수록 사용분야(업종)가 밀집하는 패턴을 보이고 있음<sup>7)</sup> (그림 3-34 참고)
- 따라서, 사용분야(업종)는 상권 발달 지역에 높게 분포되어 있으므로, 신규 사용분야(업종)은 지역 상권 중심으로 유치가 필요한 것을 알 수 있음
- 사용분야(업종)의 위치를 세분화하여 살펴보기 위해 분포 분석을 진행하였으며, 그림 3-35와 같이 문화일반 사용분야(업종) 위치 분석은 시군구 단위 문화일반 사용분야(업종) 수와 시군구 단위 사용 카드수를 비교하였음

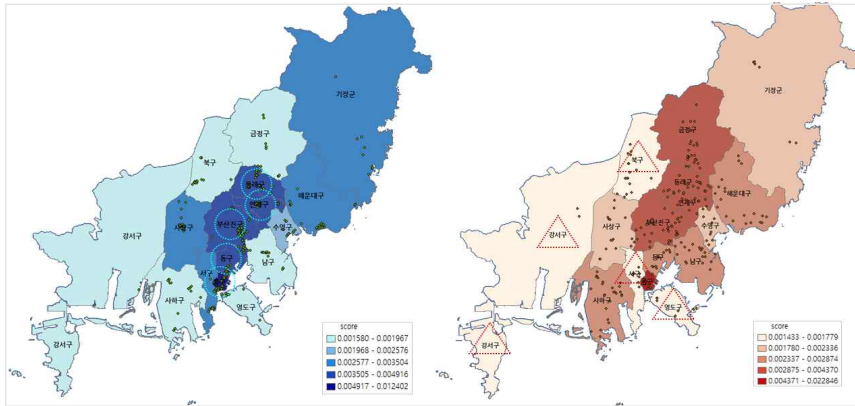
7) 그림 3-34에서 유동인구는 과란색이 진할수록 유동인구가 높으며, 카드매출액에서는 빨간색이 진할수록 카드매출이 높음

[그림 3-35] 문화일반 사용분야(업종) 분포 분석



- 그 결과 문화일반 사용분야(업종)의 분포가 상대적으로 많이 분포된 지역은 부산 중구, 부산진구, 동래구, 해운대구, 기장군으로 나타났으며, 사용된 카드 수 대비 사용분야(업종)가 상대적으로 부족한 지역은 강서구, 금정구, 연제구, 영도구로 도출되었음
- 강서구, 금정구, 연제구의 경우 타지역에 비해 상대적으로 사용분야(업종)가 낮은 지역이므로 부산시 차원에서 사용분야(업종) 확대를 위한 홍보, 지원이 필요함
- 또한 그림 3-36와 같이 숙박 사용분야(업종) 분석을 위해 시군구 단위 숙박 사용분야(업종) 수와 시군구 단위 사용 카드수를 비교하였고, 도서 사용분야(업종) 분석을 위해 시군구 단위 도서 사용분야(업종) 수와 시군구 단위 사용 카드수를 비교하였음
- 숙박 사용분야(업종) 분석의 경우 사용된 카드 수 대비 사용분야(업종)가 상대적으로 많이 분포된 지역은 부산 중구, 동구, 부산진구, 연제구, 동래구로 나타났고, 도서 사용분야(업종) 분석의 경우 상대적으로 사용된 카드 수 대비 사용분야(업종)가 부족한 지역은 부산 강서구, 영동구, 북구, 서구로 도출되었음

[그림 3-36] (좌)숙박 사용분야(업종) 분석, (우) 도서 사용분야(업종) 분석



- 사용된 카드수 대비 사용분야(업종) 현황, 위치 등을 비교분석함에 따라 사용분야(업종)를 확대 위치할 곳과 홍보가 추가적으로 필요한 곳을 도출할 수 있었음
- 이는 단순히 사용분야(업종)를 확대하기 보다는 지역적 특성 및 카드사용수, 수혜자 현황을 고려하여 확대 및 홍보 지원하여함

## □ 이상결제내역 패턴 분석

- 現 정부의 강력한 보조금 부정수급 근절을 목표로 보조금 부정수급 관리 제도와 시스템 정비, 부정수급에 대한 중첩적인 점검·감시체계 구축, 부정수급 문제에 대한 국민인식 제고를 목표로 하고 있음
- 이에 국가 보조금(66.9조 원) 수급의 공정성 강화가 필요하고, 보조금 부정수급의 관리제도 개선과 3중 점검·감시 체계 구축, 부정수급 점검 강화, 신고활성화 및 국민의식 제고의 중점 추진 과제에 맞춰 이상결제내역 모니터링 분석 방향을 설정함

[그림 3-37] 이상결제내역 모니터링 분석 방향 설정



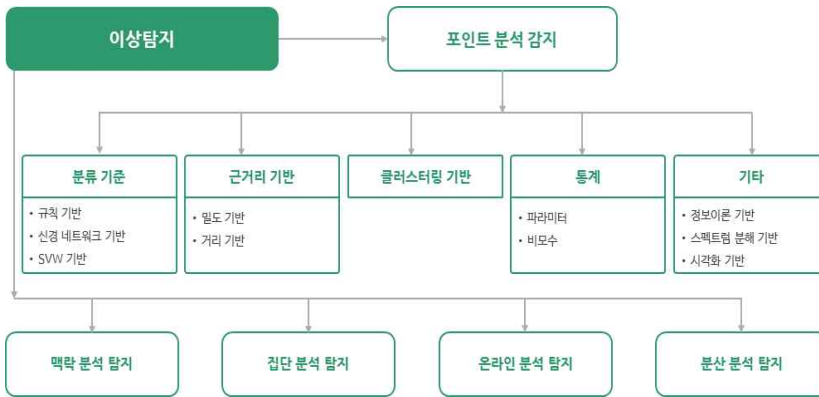
- 분석 방향 설정을 통해 이상결제내역 모니터링을 위한 패턴 분석과 이를 모델링한 예방적 모니터링 알고리즘을 개발함
- **Anomaly Detection 알고리즘을 활용한 이상결제내역 패턴 분석 알고리즘 개발**
  - Anomaly Detection 알고리즘을 활용하여 통합문화이용권(문화누리카드)의 이상결제내역 패턴 분석 방법론을 설계함

[그림 3-38] Anomaly Detection 알고리즘

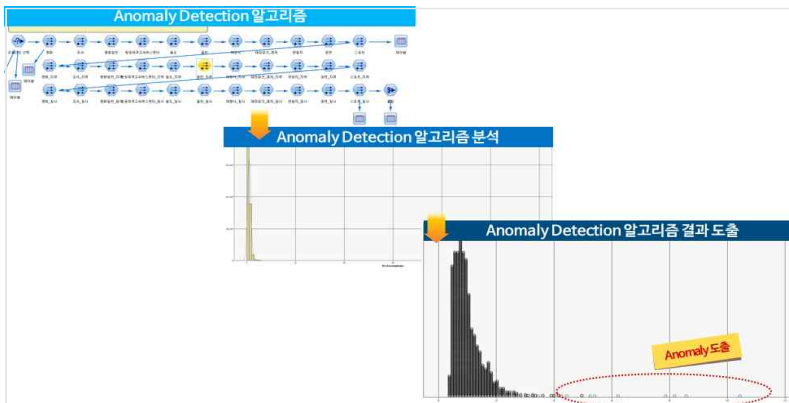


- 분류/클러스터링 분석 진행에 이상치 탐지 알고리즘을 적용하여 개인별/사용분야(업종)별 이상치 패턴을 살펴보았음
- 카드 사용 시간, 장소, 횟수 등 카드사용내역, 수혜자, 발급자 위치, 동일카드의 같은 시간, 다른 지역 사용 등 카드사용내역과 수혜자와의 관계 및 패턴 분석을 통한 이상치 패턴 분석 도출
- 통합문화이용권(문화누리카드)의 점검·감시체계를 구축하고, 이상 결제 문제를 해결하기 위해 지속적인 이상결제에 대한 패턴을 분석 할 필요가 있음

[그림 3-39] 이상결제내역 모니터링 방법론 (1)



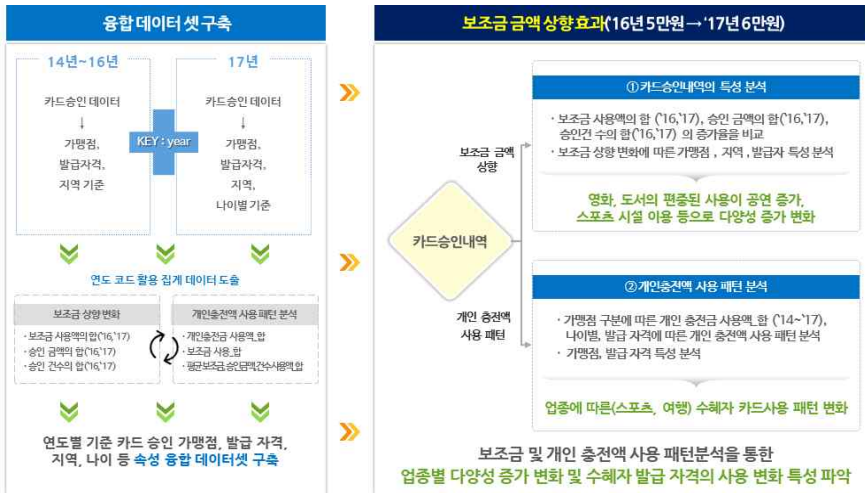
[그림 3-40] 이상결제내역 모니터링 방법론 (2)



## □ 보조금 금액 상향 효과

- 보조금 금액 상향 효과를 분석하기 위해 '14년 ~ '16년 카드사용 데이터를 사용분야(업종), 발급자격, 지역 기준에 따라 연계하고, '17년 카드사용 데이터 역시 사용분야(업종), 발급자격, 지역, 나이별 기준으로 정리하여 두 개의 데이터를 Year의 KEY 값으로 연도코드를 활용하여 집계 데이터를 도출함
- '16년과 '17년의 보조금 사용액 합, '16년과 '17년의 카드사용 금액의 합, '16년과 '17년의 카드사용 건수의 합인 보조금 상향의 변화와 개인 충전금 사용액의 합, 보조금 사용의 합, 평균 보조금카드사용 금액건 수 사용액의 합인 개인 충전액 사용 패턴 분석을 통해 연도별 기준 카드 사용분야(업종), 발급 자격, 지역, 나이 등 속성 융합 데이터셋을 구축하였음
- 카드 사용 내역을 사용건수('16년, '17년), 사용 금액('16년, '17년), 평균 1건당 사용 금액('16년, '17년)의 증가율을 비교하고 보조금 상향 효과에 따른 사용분야(업종), 지역, 발급자 특성을 분석하는 카드사용내역의 특성분석을 진행함
- 또한 사용분야(업종) 구분에 따른 개인 충전금 사용액('14년 ~ '17년)과 연령별, 발급 자격에 따른 개인충전금 사용 현황을 분석하고, 사용분야(업종)·발급 자격 특성 분석을 통해 개인충전금 사용 패턴을 분석함
- 카드승인내역에 대한 특성분석을 한 결과 기존 소액규모의 영화, 도서로 편중되었던 카드사용이 공연, 스포츠 시설 사용 등으로 가격대가 높은 사용분야(업종)으로 폭넓게 증가할 것으로 예측되며 스포츠와 여행 사용분야(업종)에 대한 수혜자 카드 사용 패턴 변화도 감지됨
- 보조금 및 개인충전금 사용 패턴분석을 통한 사용분야(업종)의 다양성 증가 변화 및 수혜자 발급 자격의 사용 변화의 특성을 파악함
- 개인충전금 사용 패턴을 '16년도 '17년도를 비교분석한 결과, 영화, 도서에서 가장 많은 소비가 발생했고, 해당 년도의 변화 수치는 미미했음

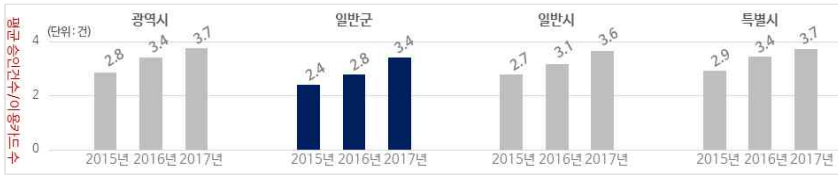
[그림 3-41] 보조금 금액 상향효과 분석 방법론



## □ 보조금 금액 상향 효과 빅데이터 분석

- 먼저, 보조금 금액 상향 효과를 살펴보기 위해 '15년 ~ '17년 광역시, 일반군, 일반시, 특별시 4개 크게 구분하여 ① 사용카드수 대비 평균 카드 사용건수(한 카드 당 평균사용건수), ② 사용카드수 대비 평균 사용금액(한 카드당 평균 사용 금액), ③평균 1건당 사용금액을 비교하였음
- 첫 번째 평균사용건수/사용카드수의 '16년에서 '17년도 증가율을 살펴보면, 특별시와 광역시는 8%로 증가하였고, 일반시는 16%, 일반군은 21% 증가한 것으로 일반군의 사용건수가 가장 많이 증가한 것을 알 수 있음 (그림 3-42 참고)
- 이는 카드 사용 건수가 증가한다는 것은 16년에서 17년 보조금의 향상이 일반군에서 효과가 있는 것으로 나타났으며, 지역적 편차가 보조금 향상에 따라 해소되고 있음

[그림 3-42] 평균 사용 건수 / 사용카드 수(한 카드 당 평균사용건수)



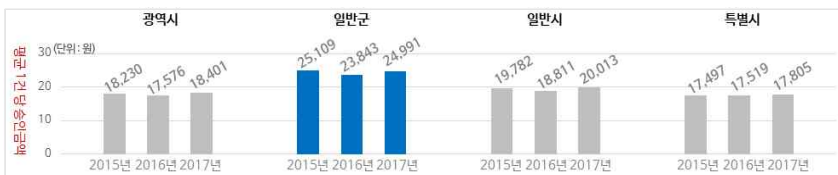
- 두 번째 평균 사용금액/사용카드 수<sup>8)</sup>의 경우 특별시와 광역시에서 각각 평균 카드사용 금액이 '17년도에 62,977원, 62,450원으로 높게 나타났고, 일반시는 62,146원, 일반군의 경우 61,210원으로 유사한 패턴을 보임 (그림 3-43 참고)
- 이는 16년에서 17년도 보조금 향상에 따라 특별시는 12% 증가, 광역시는 14%증가, 일반군, 일반시는 약 18% 증가하는 것으로 나타났음

[그림 3-43] 평균 사용금액 / 사용카드 수(한 카드당 평균 사용 금액)



- 평균 1건 당 카드사용 금액의 경우 '17년 일반군이 24,991원으로 가장 높게 나타났으며, 일반시 20,013원, 광역시 18,401원, 특별시 17,805원 순으로 나타났음 (그림 3-44 참고)
- 16년에서 17년도 보조금 향상에 따라 평균 1건당 카드사용 금액은 광역시 4%, 일반군 4%, 일반시 6%, 특별시 1% 증가 효율이 있는 것으로 나타남

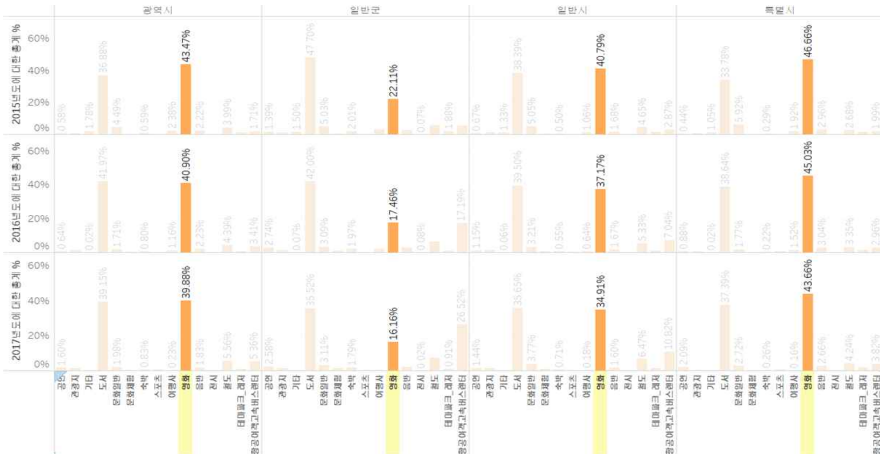
[그림 3-44] 평균 1건 당 카드사용 금액



8) 한 카드당 평균 사용 금액은 보조금, 충전금 등 합산금액을 모두 포함해서 카드 사용 횟수와 상관없이 한 카드당 평균 사용 금액을 나타는 것이며, 평균 1건 당 카드사용 금액은 한 카드당 평균 1회 사용내역을 분석 데이터임

- 즉, 보조금 상향(5만원('16년)→6만원('17년))에 따라 일반군은 '16년 대비 '17년에는 평균 카드사용건 수는 약 21% 증가, 평균 카드사용금액은 18%, 평균 1건당 사용금액 4% 증가 효과가 있는 것으로 나타났으며, 타 광역시, 일반시, 특별시 보다 평균 증가율이 4~5%정도 증가율이 높은 것으로 나타나 지역적 편차가 보조금 향상에 따라 해소되고 있음
- 사용분야(업종)에 따른 카드 사용내역을 지역별, 연도별로 살펴보면(그림 3-45), 영화는 연도가 증가할수록('15년 → '17년) 전체적으로 감소하는 추세이고 지역별에서도 전체적으로 감소하고 있다는 것을 알 수 있음
- 일반군은 상대적으로 영화의 사용이 타 지역에 비해 사용이 적은 것으로 나타났으며, 항공여객 고속버스는 '17년 기준 광역시는 5.4%, 일반군은 26.5%, 일반시는 10.8%, 특별시는 3.8%로 일반군에서 카드사용이 가장 높은 것으로 나타남. 또한, 숙박도 상대적으로 높은 비중을 차지하는 것을 알 수 있음

[그림 3-45] 사용분야(업종)에 따른 카드사용내역 특성분석(지역별, 연도별)

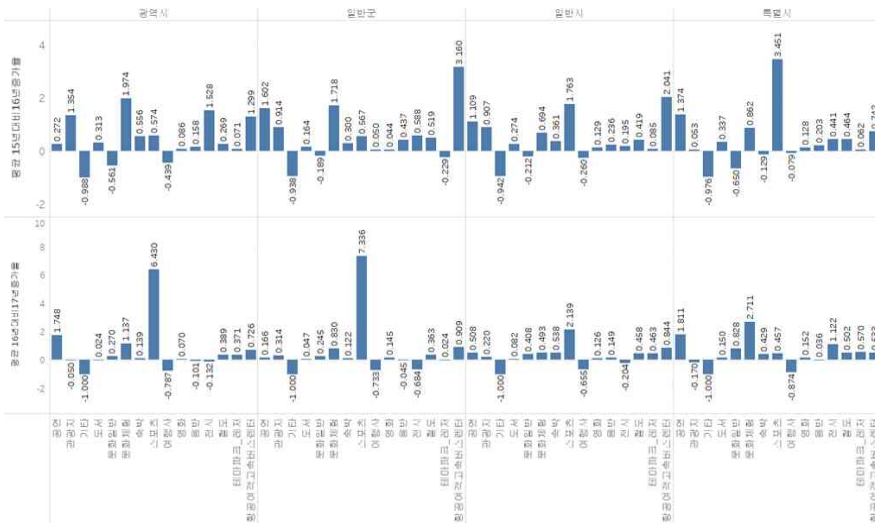


- 사용분야(업종)별 보조금 상향('17년)에 따른 증가율 변화는 스포츠<sup>9)</sup>, 철도, 항공여객 고속버스 등 광역시, 일반군, 일반시에서 전체적으로 폭 넓게 증가하고 있으며, 전사영화는 상대적으로 특별시와 광역시가 높게 증가한 것을 알 수 있음 (그림 3-46 참고)

9) '17년부터 체육시설(볼링, 탁구장, 수영장, 체력 단련장 등) 활용 가능함. 이전에는 4대 스포츠(축구, 농구, 야구, 배구)관람 및 운동용품에서만 사용이 가능하였음

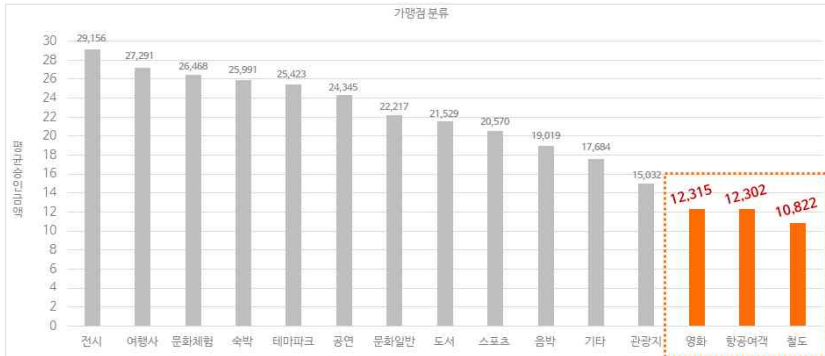
- 항공여객 고속버스는 일반시와 군 단위에서 증가하고, 문화체험 및 문화일반은 특별시가 타 지역에 비해 2배 이상 증가율이 높고 도서는 변화가 없으며, 특별시를 제외한 다른 지역에서는 전시분야 사용 비중은 감소하는 것을 알 수 있음

[그림 3-46] 사용분야(업종)에 따른 카드사용내역 특성분석(연도별 증가율)



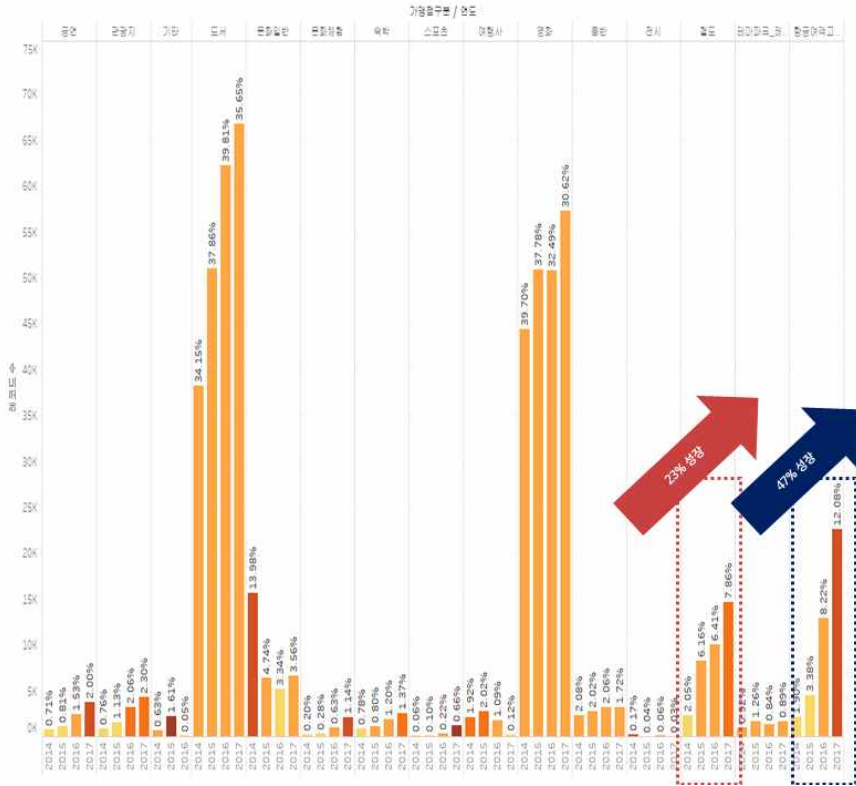
- 사용분야(업종)별 평균 사용 금액은 전시 29,156원, 여행사 27,291원, 문화체험 26,468원으로 높은 평균 사용금액을 보였고, 영화 12,315원, 항공여객 고속버스 렌터카 12,302원, 철도 10,822원으로 낮은 평균 금액을 보였음 (그림 3-47 참고)
- 사용분야(업종)별 평균 카드사용 금액은 사용분야(업종)의 이용단가를 알 수 있는 지표임. 이는 보조금이 5만원 일 경우 1회 사용비용이 적은 영화, 항공여객, 철도, 도서 등을 사용하는 경우가 많았음을 알 수 있음
- 하지만 '17년도 보조금이 6만원으로 상향되면서 1회 사용단가가 영화나 철도보다 2배 높은 '테마파크', '공연', '문화일반' 등 사용이 용이 해졌다는 것을 알 수 있음
- 보조금 상향은 도서, 영화의 유통현상을 완화시킬 뿐 아니라 다양한 사용분야(업종)의 사용 확대의 효과를 가져옴

[그림 3-47] 사용분야(업종)별 평균 카드사용 금액



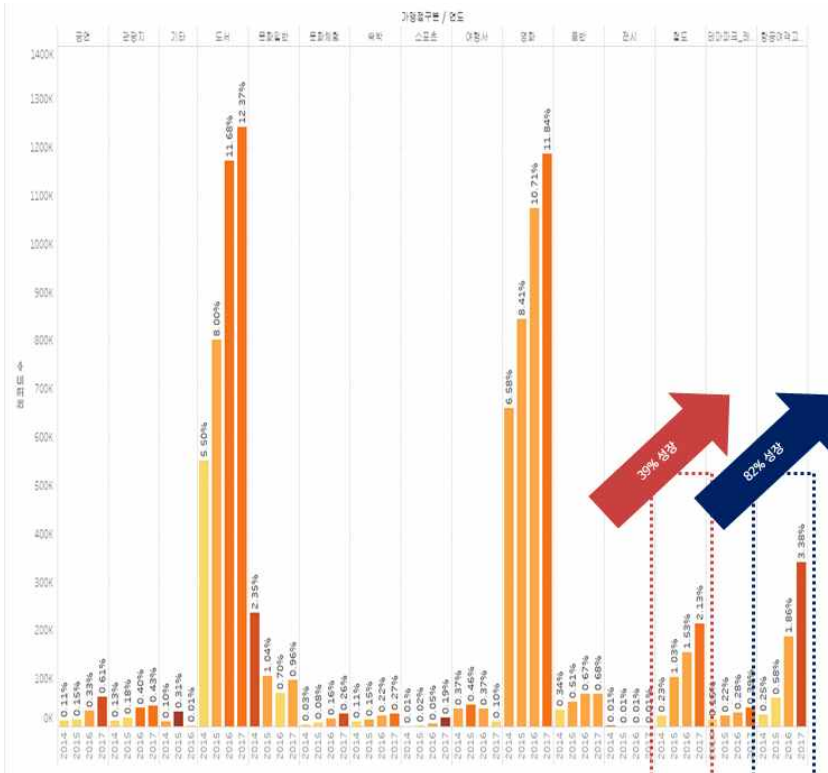
- 보조금 금액 상향 효과는 계층별 카드 사용내역을 통해 '14년 ~ '17년 장애인 계층의 사용분야(업종) 구분을 살펴보면 '14년부터 '17년까지 꾸준히 '도서'와 '영화' 분야에서 소비가 이루어진 것을 알 수 있음 (그림 3-48참고)
- 장애인 계층의 사용분야(업종) 사용현황을 살펴보면 도서의 경우 전체 건수는 늘었지만 전체 사용분야(업종) 중에 사용 비중이 '16년에는 39.8%에서 '17년은 35.7%로 4%감소한 것으로 나타났으며, 영화도 '16년 32.5%에서 '17년 30.6%로 약 2% 감소한 것으로 나타남
- 이는 통합문화이용권(문화누리카드)이 지속적으로 고민해오던 '도서'와 '영화' 쏠림현상이 '16년 5만원이었던 것이 '17년에 6만원으로 상향 조정되면서 '17년에 완화된 것으로 해석됨
- 그림 3-48을 살펴보면 장애인 계층은 철도 사용분야(업종) 이용의 경우 '16년을 기준으로 꾸준히 상승하여 '16년 대비 '17년은 23% 성장한 것을 알 수 있었으며, 항공여객 사용분야(업종)이용 역시 '16년을 기준으로 꾸준히 상승하여 '16년 대비 '17년은 47%가 성장하였음
- 즉, '도서'와 '영화'의 쏠림현상이 보조금 상향으로 인해 사용편의시설인 '철도', '항공여객'으로 이동하고 있으며, 그 외에도 스포츠 사용분야(업종)는 300%정도 증가하였고, 공연 사용분야(업종)는 30%정도 증가한 것으로 나타남

[그림 3-48] 장애인 계층 연도별 소비 사용분야(업종) 분포(2014~2017)



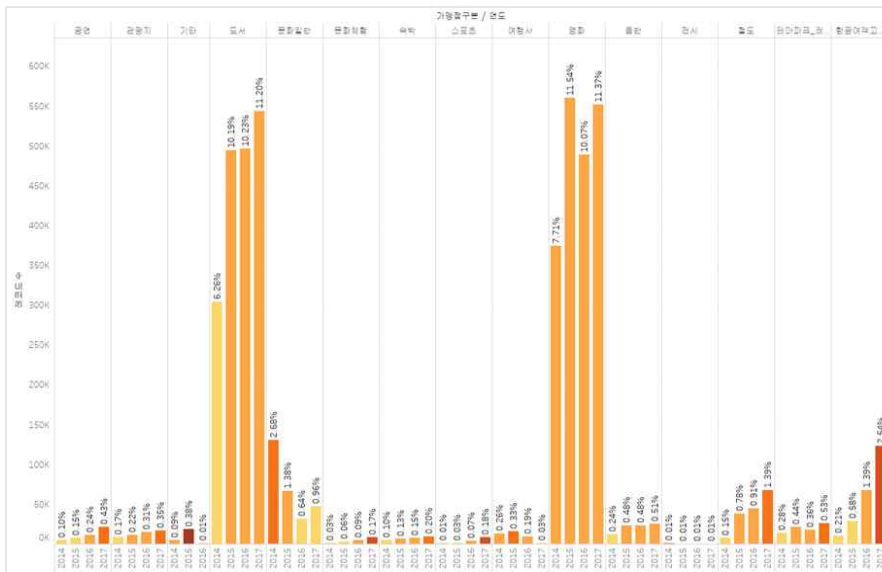
- 기초 수급자 계층의 경우 '14년 ~ '17년 사용분야(업종) 이용현황을 살펴보면 '14년부터 '17년까지 꾸준히 '도서'와 '영화'분야에 많은 소비를 한 것을 알 수 있음
- 기초 수급자 계층은 철도 사용분야(업종)의 경우 '16년을 기준으로 꾸준히 상승하여 '16년도 대비 '17년도 39% 성장한 것을 알 수 있었으며, 항공여객 사용분야(업종) 이용 역시 '16년을 기준으로 꾸준히 상승하여 '16년도 대비 '17년도 82%가 성장하였음(그림 3-49 참고)
- 이는 앞서 장애인 계층에서 분석한 것과 같이 보조금 상향(5만원('16년)→6만원('17년))에 따라 풀림현상이 완화되고 사용분야(업종) 사용 행태가 변화되는 효과가 있는 것으로 해석됨

[그림 3-49] 기초 수급자 계층의 연도별 소비 사용분야(업종) 분포(2014~2017)



- 차상위 계층의 경우 장애인 계층과 기초 수급자 계층과 같이 ‘도서’와 ‘영화’에 많은 비중을 차지하고 있으나, 연도별로 대동소이한 것으로 나타남(그림 3-50 참고)
- 차상위 계층도 철도와 항공여객은 '14년~'17년 소비가 증가하고 있으며, 철도는 '16년 대비 '17년 에 50% 증가, 항공여객은 82.7%증가한 것으로 나타났음

[그림 3-50] 차상위 계층의 연도별 소비 사용분야(업종) 분포(2014~2017)



## 제4장

---

### 결론 및 제언

제1절 : 결론

#### 1. 연구의 결론

제2절 : 제언

#### 1. 정책 활용방안 및 제언

## 제1절 결론

## 1. 연구의 결론

- 본 연구는 통합문화이용권(문화누리카드) 수혜자의 지속적 문화향유를 위한 주제별 빅데이터 분석 방법론을 설정하고, 체계적이고 과학적인 빅데이터 기술을 적용하여 분석하고자 함
- 이 연구의 목적은 통합문화이용권(문화누리카드) 빅데이터를 활용한 문화소외계층 문화향유실태분석 및 수혜자 사용분야(업종)별 맞춤형 서비스와 이상결재내역 모니터링 알고리즘 ·보조금 상향 효과 분석을 통해 개선방안을 제시하는 연구로써 구체적으로 다음과 같음

1	지역별 특성을 반영한 사용분야(업종) 확대 및 실수혜자 사용률 증대 방안 제언
	(수혜자) 카드사용 중심의 데이터를 수혜자 중심의 카드 데이터로 전환 (사용분야(업종)) 전국 사업체 총 조사 데이터를 활용한 시군구별 사용분야(업종) 비교
2	부정수급방지를 위한 예방적 모니터링 알고리즘 적용 방안 제언
	Anomaly Detection (이상치 분석) 알고리즘 모델링
3	예산 증액 등 수혜자 맞춤형 서비스 제공을 위한 의사결정 및 정책 수립 제언
	보금 금액 상향('16년 5만원 → '17년 6만원) 효과 분석 모델링

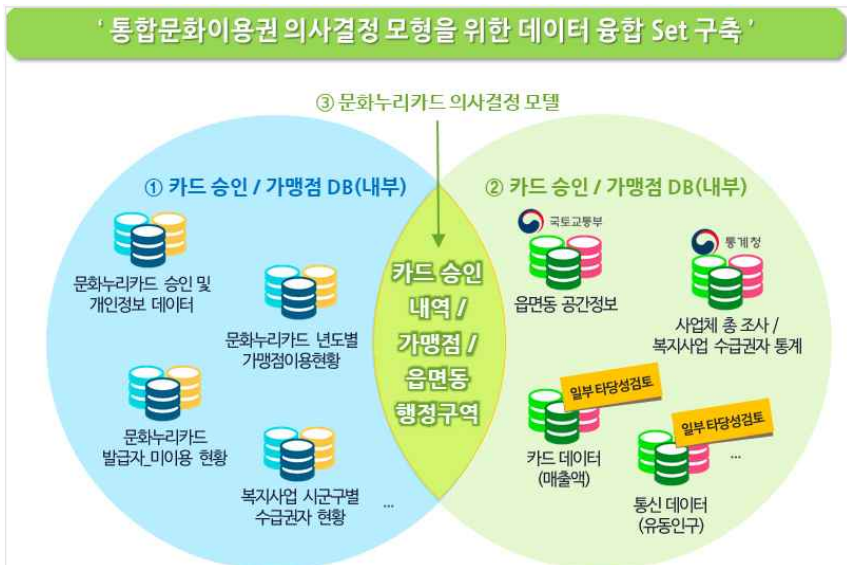
- 주요 추진한 내용은 다음과 같음

- 통합문화이용권(문화누리카드) 의사결정 빅데이터 구축

- 첫째로, 내부 데이터인 한국문화예술위원회의 문화누리카드 사용 및 개인정보 데이터, 문화누리카드 연도별 사용분야(업종) 사용현황, 문화누리카드 발급자 미사용 현황, 국토교통부 행정동 경계, 통신사 데이터, 카드사 데이터 등 외부 필요 데이터 수집 및 구축하여 융합 관리할 수 있는 DB를 구축함

- 둘째로 알고리즘을 적용한 맞춤형 서비스를 위해 데이터를 활용·융합하여 예측 결과에 대한 현황을 한눈에 파악함으로써 통합문화이용권(문화누리카드)의 지역별, 수혜자별, 사용분야(업종)별 요인과 특성을 분석하고, 이에 따른 정책적 의사결정을 제시함

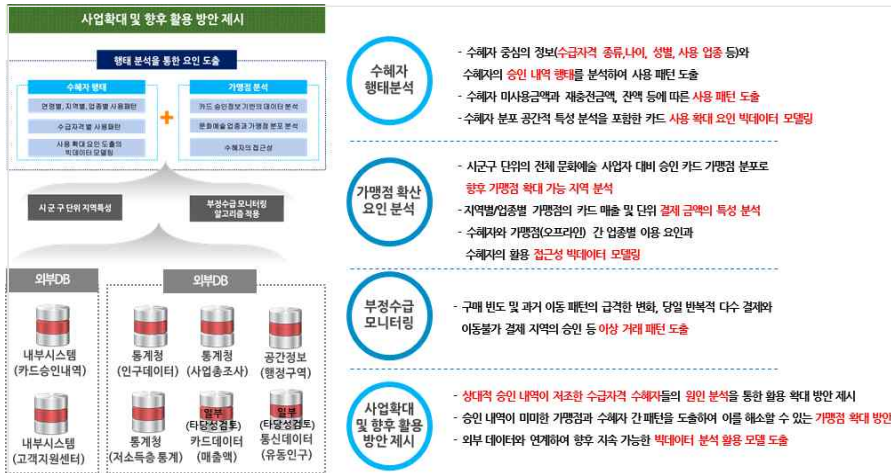
[그림 4-1] 융합 데이터셋 구축



#### — 통합문화이용권(문화누리카드) 의사결정 빅데이터 분석

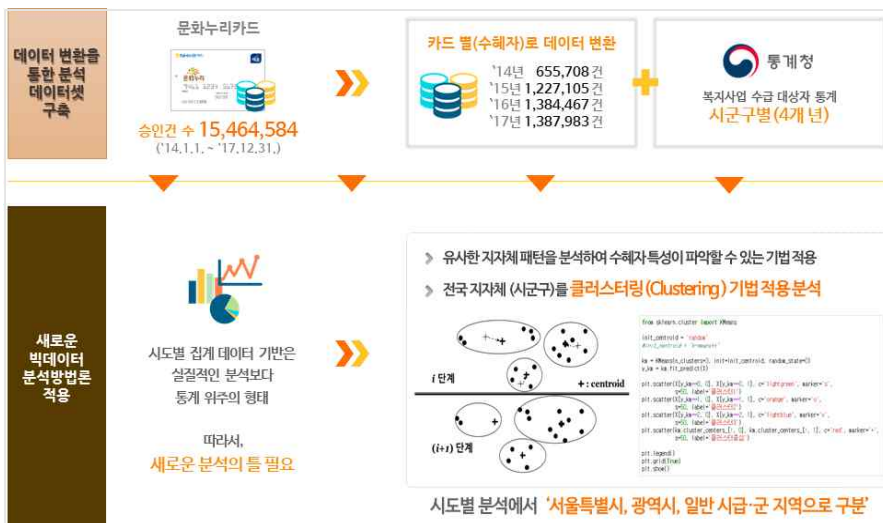
- 데이터기반의 카드사용내역과 수혜자 유형을 분석하여 보조금 상향(5만원 → 6만원 지원액의 증가)등의 효과를 객관적으로 분석하고, 이를 바탕으로 전략적 사업 확대를 위한 방향을 제시함
- 지역별 특성을 반영한 사용분야(업종) 확대 및 실수혜자 사용률 증대 방안 제언
- 부정수급 방지를 위한 예방적 모니터링 알고리즘 적용 방안 제언
- 예산 증액 등 수혜자 맞춤형 서비스 제공을 위한 의사결정 및 정책 수립 제언
- 향후 통합문화이용권(문화누리카드) 사업 활성화 방안 및 효과성 제고를 위한 빅데이터 활용방안 및 개선과제 제언

[그림 4-2] 주제별 의사결정 표준분석모델 개념 정의



- 위와 같은 표준분석모델링의 개념을 바탕으로 수혜자, 사용분야(업종), 이상결제내역 모니터링, 보조금 상향 효과 분석으로 각각 나눠 분석을 진행하고 주요 인자를 도출하여 통합문화이용권(문화누리카드)의 사용내역을 파악하고 향후 활성화 방안을 도출함

[그림 4-3] 수혜자 빅데이터 분석 방법론



- 또한, 수혜자의 지속적 문화향유를 위하여 빅데이터 분석을 통해 수혜자 맞춤형 서비스를 제공할 수 있는 기반을 마련함
- 수혜자 분석의 경우 계층별, 계층그룹의 연령대별, 사용분야(업종) 사용현황별, 사용 선호도, 전체 연령대 분포, 수혜자 대비 발급자 수 비교 등 다각적 검토를 통해 분석함
- 또한, 문화누리카드 문의사항 분석(문자 분석)을 통해 11가지 질문 유형으로 분류 한 후 질문 유형별 문화누리카드의 현황을 검토하여 수혜자들이 구체적으로 원하는 방향성에 대해 고려하였음

[그림 4-4] 수혜자 분석 결론

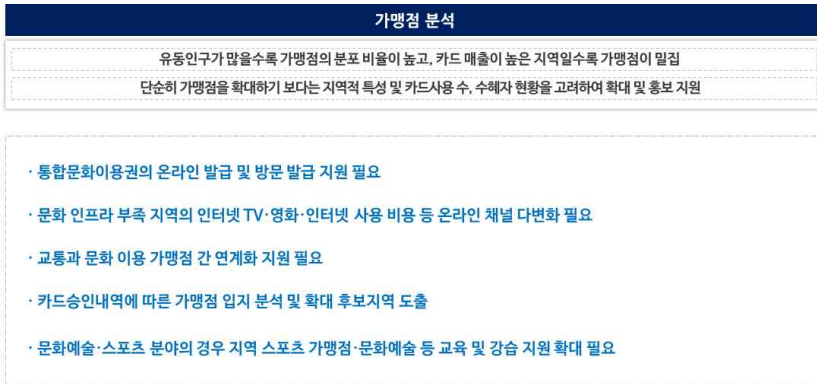
수혜자 분석	
차상위-장애인, 차상위-일반의 경우 거주지의 동일지역 소비보다 타지역에서 소비하는 패턴이 많음	
고령자, 장애인의 경우 철도·항공여객 등 생활편의 사용분야(업종) 수요가 증가함	
수혜자의 발급카드 중 카드 사용건 수는 일반군이 타광역시, 일반시보다 4%증가율을 보임	
지자체 재정자립도와 발급 수를 통해 재정자립도가 낮은 지역은 발급 수가 낮은 특성이 도출됨	

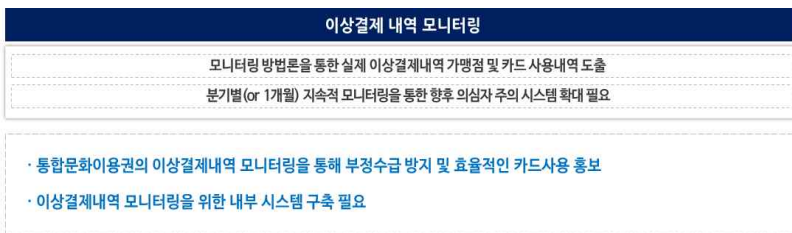
· 농촌 및 낙후지역 공공형 택시 (100원 택시), 장애인 택시 등 교통 이용 지원 확대 필요
· 고령자·장애인의 온라인 사용 방법 간편화 및 카드발급을 위한 방문 지원 필요
· 거동이 불편한 장애인, 고령자의 경우 동반자 무료 이용을 통한 문화 향유 확대 필요
· 문화예술복지사 등의 동반 이용 및 문화를 같이 공유할 수 있는 동반자 매칭 필요

- 스포츠 관련 사용분야(업종)은 다른 사용분야(업종)과 다르게 '14년 4대 스포츠 관람가능, '16년 체육용품점 허용, '17년 체육시설 허용 등 '많은 변화가 있었기 때문에 사용분야(업종) 확대에 따른 영향을 분석하기 위하여 대상으로 선정함
- 이상결제 패턴의 경우 Anomaly Detection 알고리즘을 활용하여 이상결제패턴 알고리즘을 개발하였음
- 동시 결제가 가능한 사용분야(업종)는 철도와 영화 등 전국적 사용분야(업종)를 갖고 있는 사용분야(업종)이며, 이 사용분야(업종) 들은 분 단위 결제가 가능함에 따라, 하나의 사용분야(업종)에서 분 단위 여러 건 카드결제가 일어난 곳을 추출하고, 이중 특이 사용분야(업종)를 도출할 수 있음

[그림 4-5] 사용분야(업종) 분석 결론



[그림 4-6] 이상결제 내역 모니터링 분석 결론



- 보조금 상향 효과에 따른 사용분야(업종), 지역, 카드 발급자의 특징을 분석하기 위해 카드 사용내역의 분석을 진행하였음
- 또한, 사용분야(업종) 구분에 따른 개인충전금 사용액의 합('14년 ~ '17년)과 연령별, 발급 자격에 따른 개인충전금 사용 패턴을 분석하고, 사용분야(업종)·발급 자격 특성 분석을 통해 개인충전금 사용 패턴을 분석하여 사용분야(업종) 다양성 증가 변화 및 수혜자 발급 자격의 사용 변화의 특징을 도출하였음
- 보조금 상향(5만원('16년)→6만원('17년))에 따라 일반군은 '16년 대비 '17년에는 평균 승인 건수는 약 21% 증가, 평균 승인금액은 18%, 평균 1건당 승인금액 4% 증가 효과가 있는 것으로 나타남
- 이에 따라 지역적 쏠림현상이 완화되고 사용분야(업종) 사용 행태가 변화되고, 지역적 편차가 해소되고 있음

[그림 4-7] 보조금 금액 상향 효과 분석 결론

보조금 금액 상향 효과
보조금 상향(5만원(16년)→6만원(17년))따라 영화·도서에서 스포츠·문화체험으로 비율 증가
상대적 이용 단가가 낮은 업종에서 높은 이용 단가 업종으로 변화
‘18년 7만원으로 1만원 상향조정에 따른 업종 이용 다양성 증가 기대

· 스마트 폰을 이용한 App장치 결제 지원 필요
· SNS 온라인 콘텐츠, 오프라인 문화 콘텐츠의 소액 결제 등 구매 확대 필요
· 스포츠 클럽, 공공체육시설 지원 및 이용 지원을 통해 콘텐츠화, 환경 지원 필요
· 지역축제 정보와 연계한 식당·숙박 등 다양화 필요
· 스포츠, 문화체육, 교통 등 생활편의 시설 지원 확대

- 이를 통해 ①보조금 상향효과와 ②쏟림현상 완화를 주요 사업의 성과로 제시할 수 있음

#### - 보조금 상향 효과

- 지역별 특성의 경우 카드별 평균 카드사용 건수는 특별시가 광역시, 일반시보다 상대적으로 높게 나타났지만, '16년에서 '17년 증가율은 일반군이 가장 높은 것으로 도출됨
- 카드별 사용 금액은 지역별 유사한 패턴을 보이는데 상대적으로 일반군이 사용 빈도가 낮지만 카드사용 금액이 높은 것으로 도출되어, 지역적 쏠림현상이 완화되고 있음을 알 수 있음
- 또한, '도서'와 '영화'의 쏠림현상이 보조금 상향으로 인해 사용편의시설인 '철도', '항공여객'으로 변화되고 있으며, 그 외에도 스포츠 사용, 공연 등도 증가한 것으로 나타남

#### - 쏠림현상 완화

- 사용분야(업종) 위치의 문의사항 단어 빈도 분석을 통해 비중이 많았던 영화·도서 외에도 수영장·티켓·렌터카 등의 다양한 키워드가 도출됨에 따라 향후 증가할 사용분야(업종)를 파악하였고, 체육·운동·스포츠 키워드를 통해 체육에 관한 수요가 높음을 알 수 있었음

- 또한, 테마파크·하남 스타필드·오션월드 키워드 도출을 통해 특정 지점 사용처를 문의하거나, 고속버스·버스·자전거·기차 키워드를 통해 이동수단에 관한 수혜자가 있음을 도출하여 앞으로 더 많은 분야에서 사용이 될 가능성이 있음을 확인하였음
- 장애인 계층의 사용분야(업종) 사용현황을 살펴보면 도서의 경우 전체 사용분야(업종) 중에 사용 비중이 '16년에는 39.8%에서 '17년은 35.7%로 사용분야(업종) 사용 비중은 4%감소한 것으로 나타났으며, 영화도 '16년 32.5%에서 '17년' 30.6%로 약 2% 감소한 것으로 나타남
- 장애인 계층은 철도의 경우 '16년을 기준으로 꾸준히 상승하여 '16년 대비 '17년은 23% 성장한 것을 알 수 있었으며, 항공여객 역시 '16년을 기준으로 꾸준히 상승하여 '16년 대비 '17년 47%가 성장하였음
- 즉, '도서'와 '영화'의 쏠림현상이 보조금 상향으로 인해 이용 편의시설인 '철도', '항공여객'으로 변화되고 있으며, 그 외에도 스포츠 사용은 300%정도 증가하였고, 공연은 30%정도 증가한 효과가 있는 것으로 나타남
- 기초 수급자 계층은 철도의 경우 '16년을 기준으로 꾸준히 상승하여 '16년 대비 '17년 39% 성장한 것을 알 수 있었으며, 항공여객 역시 '16년을 기준으로 꾸준히 상승하여 '16년 대비 '17년은 82%가 성장하였음
- 차상위 계층도 철도와 항공여객은 '14년~'17년 소비가 증가하고 있으며, 철도는 '16년 대비 '17년 에 50% 증가, 항공여객은 82.7%증가한 것으로 나타났음

## 제2절 제언

### 1. 정책 활용방안 및 제언

#### □ 데이터 축적 및 현행화

- 통합문화이용권(문화누리카드)의 지속적 빅데이터 분석과 활용을 위해 카드사용 및 수혜자 데이터 축적과 현행 화에 대한 전략적 접근이 필요하며, 활용빈도 높은 핵심데이터 우선 관리 및 연계 Key 관리와 업무 적용을 위한 데이터 도출이 필요함

- 한국문화예술위원회 내외부 데이터 연계 방안 수립을 수립하여 신규분석, 기존 분석의 개선을 위해 문화체육관광부, 국토교통부, 통계청 등 공공데이터와 연계한 데이터 마트를 구축하는 방안이 필요함
- 또한 연령대별, 수혜자별 맞춤형 서비스를 위해서는 통합문화이용권(문화누리카드)의 사용내역(사용 품목 내지 항목)을 확인 할 수 있는 카드내역 정보가 필요하며, 향후 사용내역을 분석하여 맞춤형 서비스 도출이 필요함
- 이는 통합문화이용권(문화누리카드)의 확대와 사업의 합리적 의사결정에 필요한 객관적 정보를 제공하는 기반이 될 수 있음

#### □ 사용분야(업종) 확대 및 실 수혜자 사용률 증대

- 통합문화이용권(문화누리카드)의 맞춤형 서비스를 제공하기 위해 고령자, 장애인 등을 위한 방문발급 등 수혜자의 공간적 제약 해소
- 카드 발급자 수의 확대를 위해 농산어촌과 낙후지역을 대상으로 방문 발급을 통한 서비스 강화
- 고령자, 장애인 계층의 경우 거동이 불편하여 문화 단절이 되는 경우가 많으므로 온라인 음악, 문화 등 사용분야(업종)를 확대 할 수 있는 방법을 제공하고, 문화 인프라 부족지역의 경우 인터넷 TV, 인터넷 영화, 인터넷 사용비용 등 온라인 채널의 다변화를 통해 실수혜자의 사용률 증가시켜야 함
- 고령자의 철도 및 항공여객의 증가율이 높아짐에 따라 문화생활과 연계한 생활편의 사용지원을 확대할 수 있는 방안을 구축해야 함
  - 공공형 택시, 장애인 택시 등 교통 사용지원을 확대하고 교통과 문화사용 사용분야(업종)를 연계하는 방안을 고려해야 함
- 스마트 폰을 사용한 영화, 도서, 음악 등 App정기 결제와 오프라인 문화를 온라인에서 구매하는 소액 결제의 방법 등 문화 콘텐츠 구매 방법 다변화와 확대를 통해 사용분야(업종)의 확대 효과 및 일반군의 사용분야(업종)의 부족, 낙후지역의 거리상 제약을 해소할 수 있음

## □ 복지 수급권자 대상 사용 확대

- 거동이 불편한 장애인과 고령자 계층의 경우 보호자, 직계 가족 등 동반자도 무료사용 할 수 있는 혜택을 제공하고, 문화시설 사용 시 이동수단이 필요할 경우 교통을 연계하여 수혜자 편의성을 증대시킴
- 장애인과 고령자 계층의 경우 공공체육시설, 생애주기별 맞춤형 스포츠 프로그램 등 다양한 생활체육시설을 동반자와 공동 이용할 수 있도록 지원 확대

## □ 업무적용

- 데이터 결과의 현황 분석 자료 및 향후 정책적 의사결정 자료로 활용되기 위해 통합문화이용권(문화누리카드) 수혜자의 변화 분석을 활용하여, 사용 빈도가 높은 지자체 또는 읍면동 단위의 세부적인 지역별 분석 확대가 필요함
- 시도별 재정자립도를 고려하여 지방비 매칭비율을 적용하고 있으나 향후 시군구별 재정자립도까지 감안한 매칭비율을 적용하는 것이 효율적인 것인 것으로 판단함

## □ 활용모델 발굴

- 카드 데이터의 사용내역 분석을 통해 복지 수급권자의 문화향유 형태 분석과 카드수혜자의 연령대별, 수급자별, 지역별의 사용형태 및 맞춤형 상세 분석이 필요함
- 또한, 사용분야(업종)의 가입 현황, 사용 현황, 폐업 현황 등 분석을 통해 기존 사용분야(업종)의 카드사용내역(목록)을 확대하고, 장애인·고령자·농산어촌 수급권자의 발급위치, 사용분야(업종) 간 공간 분석을 통한 사용분야(업종) 확대, 공간 분석을 통한 사용분야(업종) 연계 방안 다양화를 통해 접근성이동성 형태를 분석하는 등 사용분야(업종) 확대 및 향후 추진이 필요함
- 카드 미발급 대상자(계층 분석, 연령, 지역별 특성) 분석을 통한 카드 발급 확대 방안 마련과 일반군 단위에서 카드 발급 후 미사용 카드 수가 높은 편이며, 상대적으로 광역시와 특별시가 미사용 카드 수가 낮으므로 일반군 지역의 카드 발급 확대 및 미사용 카드 수 감소를 위한 고객층 세부 분석 방안도 고려를 해야 할 것임

- AR과 VR을 사용한 사용분야(업종) 홍보, 흥미·문화를 융합한 새로운 콘텐츠 생산과 나에게 맞는 (내 주변, 지역별, 연령별, 수급계층별) 사용분야(업종) 확대를 위한 추진 방안이 필요함
- 카드 사용 내역, 잔액조회, 잔액 알림 App 서비스 개발 등 AR,VR을 활용한 사용분야(업종) 위치 및 정보알림 서비스를 통해 최신 ICT 기술의 활용과 사용분야(업종) 정보를 홍보하는 등 다각도의 사용분야(업종) 활성화를 위한 추진 방법이 필요함