

경영전문석사학위 논문

Wi-Fi AP 접속 로그데이터 분석을 통한  
지역별 매장 방문 특성에 관한 연구

2023년 2월

서울과학종합대학원대학교

손상한

경영전문석사학위 논문

Wi-Fi AP 접속 로그데이터 분석을 통한  
지역별 매장 방문 특성에 관한 연구

2023년 2월

서울과학종합대학원대학교

손상한

Wi-Fi AP 접속 로그데이터 분석을 통한 지역별 매장 방문  
특성에 관한 연구

지도교수 김 보 영

이 논문을 경영학 석사 학위논문으로 제출함

2023년 2 월

서울과학종합대학원대학교

손상한

손상한의 석사 학위논문을 인준함

2023년 1 월

위원장 신 호 상 (인)

위 원 임 효 숙 (인)

위 원 김 보 영 (인)

## 초 록

스마트폰의 대중화, 5G 네트워크 기술의 발전과 더불어 다양한 형태의 대용량 데이터를 처리할 수 있는 빅데이터 관련 기술의 발달로 고객들의 실시간 위치정보의 수집 및 이를 활용한 분석이 가능하게 되었다. 특히 온라인 쇼핑몰 환경에서는 고객의 진입에서부터 최종 결제까지 이뤄지는 구매결정 전 과정에 대해 방대한 량의 데이터를 수집하고 이를 통해 고객의 온라인 매장의 이용형태를 분석하고 있다.

그러나, 21년 통계청 자료에 의하면 518조 규모의 소매 유통시장에서 온라인 매출 규모는 187조 규모로 여전히 오프라인 매장에서의 매출이 큰 비중을 차지하고 있다. 이에 반하여 온라인 매장과 같은 방대하고 다양한 종류의 고객 데이터 수집을 오프라인 매장에서 진행하기에는 어려운 환경이다. 따라서 온라인 매장에서 진행했던 것과 같은 데이터 기반의 다양한 고객 행동 분석이 힘들고 온라인 매장에서의 고객 행동 분석 대비 오프라인 매장에서의 고객 데이터 분석이 상대적으로 열세인 상황이다.

이러한 문제점을 해결하고자 최근 오프라인 매장내 설치되어 있는 Wi-Fi AP(Access Point)에서 수집된 고객 단말기 정보를 통해 이러한 문제점을 해결해보고자 하는 연구들이 진행되었다. 특히, 최근 대부분의 오프라인 매장에서 다양한 인터넷 서비스를 제공하고자 Wi-Fi AP(Access Point)를 설치하여 운영하는 곳이 많아졌고, 이를 통해 고객의 스마트폰을 통해 Wi-Fi AP에 접속한 데이터를 수집함으로써 고객들의 매장내에서의 동선 추적 및 이를 통한 구매에 영향을 미치는 요소들에 대한 연구도 진행되었다.

그러나, 그동안 진행된 연구들은 각각의 단일 매장에서의 데이터만을 대상으로 연구해왔고, 프랜차이즈 업체나 대형 유통사가 아닌 환경에서는 각 매장의 데이터 공유가 어렵고, 수집된 AP 데이터의 형태도 각 제조사 별로 상이하어 데이터 통합에도 여러 이슈들이 발생하여 여러 매장에 대한 통합적인 고객 연구는 어려웠던 사항이 있었다. 따라서 단일 매장내 고객행동 분석 중심으로 연구가 진행되어져 왔고, 좀더 거시적인 지역이나 상권중심 단위에서의 고객 행동분석은 더욱이 어려웠다.

본 논문에서는 기존 연구에서의 매장내 Wi-Fi AP에서 수집한 고객 단말기 정보가 아니라 고객들의 스마트폰에서 수집된 방문한 매장의 Wi-Fi MAC Address 접속 로그 데이터를 확보하여 이를 매장의 위치, 업종 데이터와 결합하여 시간대별, 지역별, 업종별 고객 방문 특성에 대한 차이점을 분석하였다. 총 18개월간 6억여건의 Wi-Fi 접속 데이터를 수집하였고, 이를 데이터 전처리 과정을 통해 총 186만여건의 고객의 매장 방문 데이터로 정제하여, 시간대별, 지역별, 업종별 고객들의 매장 방문 특성에 대한 차이를 각 매장단위가 아닌 지역별, 업종별로 거시적으로 바라보고 분석하여, ‘시간대별 고객방문자 수’와 ‘시간대별 매장 체류시간’이라는 속성을 통해 고객의 오프라인 방문매장 특성에 대한 지역별 유사도 및 차별적 고객 행동 특성이 보이는 지역에 대한 고찰을 제시한다.

주로 서울지역을 대상으로 각 행정구역 ‘구’단위로 분석을 진행하였고, 서울시 전체적으로 고객들은 주로 12시, 18시를 중심으로 한 식사를 목적으로 하는 업종에서 많은 고객 방문이 발생하였으며, 시준적으로는 봄(3~5월), 가을(9~11월)이 다른 계절보다 높은 매장 이용 형태를 보이는 것으로 분석되었다.

이와 더불어, 일부 지역에서는 다른 지역과는 다르게 아침식사대(8시)

에도 많은 매장 방문자 패턴이 나타났으며, 서울 외곽지역 중심으로 주로 저녁식사 시간대 집중적으로 매장 방문패턴도 발견되었다.

기존 연구에서 많이 활용되어왔던 코사인 유사도 분석을 활용하여 본 연구에서는 시간대별 방문자 수 외에 시간대별 체류시간 평균과 분산을 활용하여 지역별 고객 매장 방문 이용형태에 대한 유사도 분석의 분해성이 10% 정도 증가됨을 확인하였고, 이와 같이 보다 개선된 유사도 분석 결과를 통해 지역기반 추천 알고리즘에 적용한다면 추천 성능 개선에 긍정적인 영향을 미칠 것으로 예상된다.

키워드: 와이파이, 로그데이터, 매장, 방문특성, 방문패턴, 체류시간, 지역 유사도, 서울

# 목 차

I. 서론 .....	1
1. 연구 배경 .....	1
2. 연구 목적 .....	3
3. 논문의 구성 .....	4
II. 이론적 고찰 .....	5
1. 고객의 매장 방문 특성에 관한 분석 .....	5
1.1 데이터 기반 매장 방문 개념 및 특성 .....	5
1.2 매장 방문 특성에 관한 기존 선행 연구 .....	7
2. 고객 데이터 분석의 개념과 방법 .....	8
2.1 POI(Point of Interest) 추천에 관한 연구 .....	8
2.2 코사인 유사도(Cosine Similarity) 분석 및 활용 연구 ..	10
2.3 Wi-Fi 로그 데이터 활용 연구 .....	13

III. 연구 방법 .....	15
1. 데이터 수집 및 전처리.....	15
1.1 프리파이(Free-Fi) 앱 서비스 소개 .....	15
1.2 Wi-Fi AP 접속 로그데이터 (AP Log Data) .....	16
1.3 매장 정보 데이터 (Shop Data) .....	19
1.4 분석용 마트 데이터 (Mart Data) .....	20
1.5 최소 체류시간 기준 .....	22
2. 데이터 시각화 및 코사인 유사도 모델 설계 .....	24
IV. 분석 결과 .....	27
1. 지역별, 시간대별 방문자 수.....	28
2. 지역별, 시간대별 체류시간 .....	31
3. 지역 유사도 분석 .....	33
V. 결론 .....	38
1. 요약 및 시사점.....	38
2. 연구의 한계 및 향후 계획.....	40

참 고 문 헌 .....	42
감 사 의 글 .....	47

## 표 목 차

<표1> 코사인 유사도에 대한 파이썬 코드 예시 .....	12
<표2> AP 로그데이터 속성 .....	17
<표3> AP 로그데이터 예시 .....	18
<표4> 매장 기본 정보 .....	19
<표5> 최소 체류시간 기준에 따른 데이터셋 비교 .....	23
<표6> 매장 방문자 수 증가도에 따른 지역 .....	30

## 그림 목 차

<그림1> 코사인 유사도 개념 .....	11
<그림2> 매장 데이터 지역별 분포 현황 .....	20
<그림3> 업종 범주형 데이터 분포 현황 .....	21
<그림4> 분석용 마트 데이터셋 종합 .....	22
<그림5> 분석용 마트 데이터셋 예시 .....	22
<그림6> 최소 체류시간 기준에 따른 패턴 변화 (1분 vs. 3분) .....	24
<그림7> 입력 속성 증가에 따른 코사인 유사도 변화 개념 .....	25
<그림8> 시간대별 매장 이용 결과 테이블 예시 .....	27
<그림9> 전체 매장 방문 형태 .....	28
<그림10> 삼시 세끼형 시간대별 방문자 수 패턴 .....	29
<그림11> 저녁장사형 시간대별 방문자 수 패턴 .....	30
<그림12> 서울시 관악구 vs. 강남구 vs. 동대문구 방문자수 패턴 .....	31
<그림13> 전체 매장 평균 체류시간 및 분산 .....	32
<그림14> ‘식당-한식’업종 매장 평균 체류시간 및 분산 .....	32
<그림15> 코사인 유사도 파이썬 구현 예시 .....	33
<그림16> 유사도 편차 최대 발생지역 비교 (서대문구 vs. 강서구) .....	35
<그림17> 유사도 편차 최소 발생지역 비교 (강남구 vs. 강동구) .....	35

<그림18> 코사인 유사도 모델 별 지역간 유사도 계수.....	36
<그림19> 매장 방문 패턴 비교 (서울시 서대문구 vs. 마포구).....	37

## 제 I 장 서 론

### 제1절 연구의 배경

최근 인터넷 쇼핑과 모바일 쇼핑의 급성장으로 인해 가상공간내에서의 고객들의 구매활동 비중이 크게 확대된 것은 사실이나 아직까지는 매장내에서의 구매활동 비중이 크다는 것은 사실이다. 하여 많은 기업과 상점에서 매장의 매출 증대를 위해 고객들에 대한 행동 분석과 그들의 구매 결정 의사에 영향을 미치는 요소들에 대한 관심과 투자는 오늘날에도 지속되고 있다.

그러나 현실적으로 오프라인 매장에서의 고객 행동에 대한 데이터를 확보하기란 매우 어렵거나 상당한 비용이 수반되는 일이었다. 별도의 조사원을 동원하여 일일이 카운터 체크기와 같은 것을 이용하여 직접 방문자 수를 조사하거나, 매장에 방문한 고객을 대상으로 설문조사를 통해 관련 데이터를 확보하는 식으로 진행되어져 왔다. 혹은 포커스 그룹 인터뷰(FGI) 방식과 같이 제한된 인원을 대상으로 추적/관찰하는 방식으로 데이터를 수집하기도 한다. 이러한 데이터 수집 방식은 많은 량의 데이터를 수집하기에 막대한 시간과 비용이 필요하고, 또한 수집된 데이터에도 고객의 주관적 의도성이 많이 반영될 수 있어 데이터 객관성 확보에도 어려움이 많았다.

이러한 문제점들을 해결하고자 많은 시도들이 있었으며 최근 고도화된 IT기술의 발달과 스마트폰 대중화로 인하여 과거에는 확보하기 어려웠던 고객 관련 데이터를 손쉽게 수집할 수 있게 되었다. 대부분의 고객들이 스마트폰을 보유하고 있으며, 일상생활 중에 거의 모든 장소와 시간에 스마트폰을 소지하고 생활하고 있다. 이러한 현상으로 인해 고객의 위치 정보에 대한 데이터를 스마트폰을 통해 확보하는 것이 가능하게 되었고, 스마트폰을 통해 수집된 데이터를 통해 개인화 추천(Personal Recommendation) 알고리즘의 발전이 최근 수년 동안에 진행되었다. 즉, 고객이 관심 있어 할 만한 상품에 대한 추천을 고객 각 개인의 특성에 맞춰서 추천하는 형태의 영

업전략이 온라인 시장을 중심으로 활발히 진행되었다.

또한 여전히 매출 비중이 높은 오프라인 매장의 매출 증대를 위해 고객들의 동선 분석이 중요해짐에 따라 POI(Point of Interest)에 대한 추천 기술이 주목받고 있다. POI란, 소비자가 흥미롭거나 유용하다고 여기는 특정한 장소를 의미한다. 이와 관련된 연구로 POI추천시스템 (박소현, 박영호, 박은영, 임선영, 2018), 오프라인 쇼핑몰에서의 사용자 이동패턴 분석 연구(최영환, 이상용, 2006) 등이 진행되었다. 우선 매장의 매출을 증대하기 위해 고객 유입 극대화를 위한 전략으로 기존에 매장을 방문했던 고객들의 동선을 분석하여 이와 유사한 동선을 보이는 고객들에게 매장을 추천함으로써 고객 유입 효과를 기대하는 방식이다.

따라서 이러한 POI 추천 기술에서 유사도 및 상관관계를 분석하는 것은 매우 중요하다. 이미 많은 분야에 다양한 유사도 분석 방법을 사용한 연구들이 진행되고 있는데, 예를 들어 사이버 여론 조작 감지 응용 연구(김민재, 이상진, 2014), 이주패턴 변화 응용 연구(한이철, 2017), 커뮤니케이션 분야 활용 연구(김승환, 박진균, 한상용, 2010)등이 그 예이다.

그러나 이전에 진행되었던 POI 추천시스템 관련 연구들은 특정 데이터셋에 한정되어 연구되었다는 한계점이 있다. 스마트폰 활성화 이전에는 개인 위치정보를 수집하기가 매우 어려웠고, 스마트폰을 통한 개인 위치정보를 수집한 이후에도 개인 정보 보호 이슈와 매장간 데이터 공유에 대한 인식 부족 등으로 인해 데이터 확보 및 통합이 용이하지 않았다. 정부에서도 4차산업혁명의 근간이 될 수 있는 데이터산업 육성을 위해 ‘서울 열린 데이터 광장<sup>1)</sup>’등을 통해 공공데이터들에 대한 공유 등의 지원을 지속하고 있으나, 개인식별그룹으로 그룹화된 통계량적 데이터만 제공되는 등 실제적인 데이터 확보의 어려움이 존재하였다. 또한 수집되는 데이터 간에 표준화도 진행되어 있지 않은 상황이라 어렵게 서로 다른 매장 간의 고객 방문데이터를 확보한다고 하더라도 데이터를 결합하여 하나의 통합 데이터베이스로 구성하기가 힘

---

<sup>1)</sup> <https://data.seoul.go.kr/>

들었다. 하여 다양한 매장에 대한 데이터 확보가 어렵고 데이터의 결합성, 활용성도 낮은 수준이라 POI 추천 연구에서 많이 사용된 유사도 분석의 입력변수를 단순히 방문빈도와 같은 1차원적 입력변수로만 연구를 진행해왔다. 일부에선 단일 매장을 대상으로 하여 방문빈도나 체류시간 등의 다양한 입력변수로 매장내 고객행동분석 연구(김소진, 2016)가 진행되었으나 제한적 매장 내라는 환경적 제약을 벗어나지 못한 점이 있고 관광산업에서 빅데이터를 이용하여 관광객 유사성에 관한 연구(장운정, 정수아, 2021)는 있었으나, 거시적 관점에서의 업종별, 지역적 특성 분석에는 연구가 부족한 실정이다.

따라서 본 연구에서는 개인 스마트폰에 설치된 별도의 앱으로부터 수집된 Wi-Fi 접속 로그 데이터를 사용자 매장 방문 데이터화 하여 매장정보 데이터와 결합하여, 지역간, 업종간 유사도를 분석하며, 이들의 유사도 측정의 정확도를 향상시키고자 새로운 입력변수에 대한 제안 및 그에 따른 성능을 검증하고자 한다. 기존 Wi-Fi 접속 로그 데이터를 활용한 연구와는 다르게 본 연구에서는 Wi-Fi AP에서 수집한 개별 스마트폰 정보가 아닌, 개별 스마트폰 내에 저장되어 있는 Wi-Fi AP 접속 데이터를 활용한다. 또한 유사도 분석을 위하여 본 연구에서는 다양한 분야에서 이미 활용되고 있는 코사인 유사도(Cosine Similarity) 분석 방법을 사용하였으며, 본 연구에서 활용된 Wi-Fi 로그데이터에 대한 전처리 과정에 대해서도 간략히 소개하여 Wi-Fi AP 접속데이터라는 raw data로부터 고객 방문이력 데이터로 가공/활용하는 방안에 대해서도 검증하고자 한다.

## 제2절 연구 목적

본 논문에서는 18개월간 서울, 경기 지역의 오프라인 매장에 설치된 Wi-Fi AP에 접속한 고객 스마트폰 로그 데이터와 해당 매장의 업종, 위치 정보 등을 활용하여 다음과 같은 2가지에 대해 분석하고자 한다. 먼저 Wi-Fi 접속 로그 데이터를 활용하여

고객 방문데이터로 변환하고 그에 따라 각 지역별 고객 방문 패턴의 차이를 살펴보기로 한다. 하여 데이터 수집을 위한 부수적 제반시설 설치 과정 없이 이미 운영되고 있는 Wi-Fi 서비스의 접속 로그데이터를 활용하여 고객 방문 데이터로서 활용성을 검증하고, 이를 통해 지역별 고객 매장 이용형태에 대해 살펴볼 것이다. 다른 하나는 지역 간에 유사도 분석 때 체류시간의 특성을 새로운 입력변수로 적용하여 시간대별 방문자 수 중심의 단일 입력변수로 진행하였던 기존의 유사도 분석 연구와의 차이를 살펴본다. 이를 통해 좀더 유사도 분석의 정확성이 향상된다면 POI 추천 성능도 개선될 것이고, 거시적 관점에서의 유사 지역간 상관분석도 가능할 것이다.

따라서 본 논문의 연구 목적은 다양한 매장에서 획득한 고객의 방문데이터를 분석한 연구로써 고객 방문에 대한 별도의 센서 측정 데이터에 기반한 연구가 아닌 일반적으로 거의 모든 매장에서 사용하고 있는 Wi-Fi AP에 대한 접속 데이터를 이용하여 별도의 부가적인 센서, 설비 등의 설치 없이 고객의 매장 이용형태에 대한 분석에 적용할 수 있는 방안을 연구한다. 또한 기존에 많이 활용되고 있는 코사인 유사도 분석 기법에 새로운 입력변수를 제안함으로써 기존 연구에서 활용되던 코사인 유사도와 본 연구 간의 유사도 비교실험을 진행한다. 이를 통해 지역간 코사인 유사도 분석에 있어서 분해능을 높여 좀더 정확한 유사도 분석을 통해 협업모델 기반 추천과 같은 POI 추천 성능 개선에 기여하고자 함이다.

### 제 3 절 논문의 구성

본 논문의 구성은 다음과 같다. 2장에서는 고객 이용형태 분석이란 무엇인지 또한 고객 데이터 분석 개념과 그 방법들에 대해 간략히 소개한다. 3장에서는 본 연구에서 사용한 데이터의 수집방법과 수집된 데이터를 어떻게 가공하였는지에 대한 데이터 전처리 과정에 대해 설명한다. 또한 지역별 고객 방문 데이터의 유사도 분석을 위한 코사인 유사도 분석 방법을 소개한다. 4장에서는 고객 매장 이용형태에 따른 서

울/경기 지역 전체에 대한 특성과 특징적 패턴이 나타난 지역에 대한 개별 지역 결과를 분석한다. 5장에서는 결론을 내리고 향후 연구에 대해 소개한다

## 제 II 장 이론적 고찰

2장에서는 기존의 고객 매장 이용형태 분석 및 고객 데이터 분석에 관한 개념과 방법에 관한 연구들을 설명한다. 먼저, 1절에서는 고객의 매장 이용형태에 관한 연구들과 그 방법론에 대해 소개하고, 2절에서는 고객 데이터 분석 개념들과 다양한 활용 연구에 대해 소개한다.

### 제1절 고객의 매장 방문 특성에 관한 분석

#### 1.1 데이터 기반 매장 방문의 개념 및 특성

일반적으로 고객이 구매활동을 위해선 매장이라는 장소를 방문하게 된다. 고객은 자신이 보유한 재화와 본인이 소유하고자 하는 물건을 교환함으로써 매매라는 경제활동이 발생되고, 이러한 경제활동이 발생하는 장소를 우리는 매장이라고 한다. 따라서 여기서 말하는 매장이란 꼭 오프라인 상점을 의미하는 것만은 아니고 인터넷 쇼핑몰이라는 온라인 장소도 동일한 매매라는 경제 활동이 이루어지는 곳으로 매장에 포함할 수 있을 것이다. 하여 많은 기업에서 매출 증대를 위하여 매장내에서의 고객의 행동을 분석하거나 매장내 구매 의사 결정 단계에 대한 연구(이세진, 이정교, 2006)는 예전부터 많은 연구와 시도들이 있었다. 이러한 연구에서 고객의 구매 결정에 영향을 미치는 요소들에 대한 정의(김혜영, 송지희, 2015)와 그것들의 영향도 등을 수치화 또는 계층화하여 활용가능한 정량적 데이터로 변환하는 많은 연구(장혜

원, 2021)들을 제안하고 있다. 그러나 과거의 연구들은 고객들의 비정형적 행동을 정량적 데이터화 하는 과정에서 많은 비용과 왜곡들이 발생하기도 하였으며, 이렇게 가공된 데이터들에 대한 신뢰성을 확보하는데 부족한 점이 있다.

최근에 온라인 시장의 성장과 전체 시장의 약 20% 수준의 매출 규모 비중을 차지할 만큼 그 중요도가 증가하고, 오프라인 매장에서의 비정형 데이터 수집보다 저렴한 비용으로 손쉽게 고객들의 행동에 대한 정형 데이터 수집이 가능하다는 장점으로 온라인 매장 내에서의 고객 매장 이용형태 분석(이자혜, 2018)이 활발히 진행되었다. 특히 온라인 매장의 경우엔 고객이 어떤 과정을 통해 구매 아이템을 발견하는지, 아니면 어떤 순간에 관심 아이템에서 이탈하는지 등 구매결정과정 뿐만 아니라 네거티브 행동들에 대한 데이터도 수집할 수 있다는 장점이 있다(김다운, 2021). 특히, 구글 애널리틱과 같은 온라인 매장내에서 고객의 행동 형태에 대한 데이터 수집 솔루션의 발달과 이렇게 수집된 대용량의 데이터를 저장, 처리할 수 있는 빅데이터 기술의 발달로 인해 온라인 매장에서의 고객 행동 분석은 이미 상당한 수준으로 발전되었고 이미 많은 부분에서 활용되어 그 성과도 나타내고 있다.

온라인 매장에서 고객 행동 분석 시 주로 활용하는 지표들은 단위 기간 동안 순방문자수(Unique Visitor, UV), 고객 유지률(Retention Rate), 이탈률(Bounce Rate), 평생가치(Life Time Value, LTV), 페이지 뷰(Page View, PV), ARPU(Average Revenue Per User), 전환률(Conversion Rate, CVR), 전환비용(Conversion Cost), 고객획득비용(Customer Acquisition Cost, CAC), 광고비 대비 수익(Return On Advertising Spending, ROAS) 등 다양한 형태로 이미 운영되고 있다.

본 연구에서는 이렇게 온라인에서 활성화된 고객 분석 기법을 오프라인 매장에서 확대 적용해보고자 한다. 기존에도 이러한 시도들은 많이 있었으나 오프라인 매장에서의 고객 행동들을 데이터화 하기가 어렵고 또한 이를 수집하기에 많은 비용이 소요되어 이러한 연구가 진행이 쉽지 않았다. 하여 비록 직접적인 형태의 데이터 수집은 아니지만, 고객들이 매장에 방문하게 되었을 때 고객 단말기가 매장 Wi-Fi AP( Access Point)에 접속한 데이터로써 고객의 매장 방문여부와 머문 시간 등에

대해 빅데이터 처리를 통해 간접적으로 데이터를 확보할 수 있다. 또한 온라인 매장의 고객 행동 분석 때 사용되는 ‘단위시간당 순 방문자수’와 유사한 개념으로 ‘시간 대별 매장 방문자 수’와 온라인 매장에서 활용되었던 ‘고객 유지률’과 ‘페이지 뷰’와 같은 ‘매장 방문 고객 매장내 체류시간’이란 데이터를 활용하여 고객의 매장 방문 특성에 대해 분석하고자 한다.

## 1.2 매장 방문 특성에 관한 기존 선행 연구

온라인 쇼핑몰의 고객 행동 분석이 중요해짐에 따라 서비스 내 사용자 행동을 분석하기 위한 국내외 연구들 역시 활발히 진행되고 있다. 온라인 쇼핑몰에서 이용자의 실제 행동이 기록되어 있는 웹 로그 데이터 및 클릭스트림 데이터와 구매 데이터 등을 기반으로 행동 패턴을 분석하거나, 행동 특징 별 군집화 분석을 하는 연구 등이 진행되고 있다(박기연, 2019). 다양한 분석기법과 예측 모델을 활용한 재구매 예측 연구와 이런 재방문과 재구매 및 이탈 요인에 대한 연구(추희정, 2022)도 진행되고 있다.

이들 연구들은 기존의 설문조사 방식의 데이터 수집 방법이 아닌 IT기술 발달로 인해 고객이 온라인 쇼핑몰에 접근하는 순간부터 이탈하는 마지막 순간까지 발생했던 데이터를 활용하였다. 방문한 웹 페이지의 웹 로그 데이터, 사용자 마우스의 움직임 기록하는 클릭 스트림 데이터, 구매 관심 아이템을 저장하는 장바구니 데이터와 실제 구매/결제까지 완료한 구매데이터 등 사용자의 행동에 따라 자동적으로 발생하는 데이터를 수집함으로써 기존보다 손쉽고 비용도 크게 들지 않고 다양하고 풍부한 객관적인 데이터를 활용한 연구들이 최근 많이 진행되고 있다.

이렇게 수집된 데이터들은 수치화 하거나 계층화를 통해 통계분석이 용이한 데이터 형태로 가공하기 용이한 장점이 있다. 화장품 쇼핑몰의 고객 재방문과 재구매 및 이탈률에 대한 원인에 관한 연구(추희정, 2022)에선 이러한 데이터를 통해 고객

의 나이나 회원 등급, 지역 등은 재방문이나 이탈에는 유의미한 영향이 없음을 확인하였고, 총 구매금액, 방문시간대, 평균 접속시간, SNS 추천 등의 인자에 의해 재방문에 양의 영향을 미치는 것으로 분석되었다.

또한 최근에는 추천시스템의 성능 향상을 위해 비정형 데이터를 활용해야 한다는 필요성이 꾸준히 제기되어져 왔고, 특히 사용자의 성향이나 선호도와 같은 부분을 추천에 반영하기 위해 이러한 비정형 데이터를 활용하는 연구(최성이, 2016)들도 진행되었다. 사용자가 임의로 입력한 취미와 관심분야 등에 국한된 정형 데이터를 활용하는 것이 아니라 사용자의 최근 인터넷 사용패턴이나 최근 쇼핑물내 상품 상세페이지를 본 상품 정보 등의 비정형 데이터로부터 고객 관심사라는 정형성 데이터를 추출해내고 이런 결과를 추천 시스템에 반영함으로써 추천 성능 향상을 도모할 수 있는 방안을 제시하기도 하였다.

본 연구에서는 기존의 연구들과는 다르게 다수의 매장에 대한 데이터 분석으로 해당 지역에 대한 거시적 관점에서의 방문자 패턴을 분석하고자 한다. 대부분의 추천 시스템에 관련된 연구나 고객 행동 분석이 단일 매장 내에서 발생하는 패턴에 대한 연구가 중심이라면 본 연구는 행정 구역별 매장들에 대한 방문 패턴을 분석해 봄으로써 지역별 유사도에 관하여 연구해보고자 한다.

## 제 2 절 고객 데이터 분석의 개념과 방법

### 2.1 POI (Point of Interest) 추천에 관한 연구

온라인에서는 이미 구글 애널리틱과 같은 방문자 추적 시스템이 널리 사용되고 있으며, 센서 기술의 발달로 오프라인 매장에서도 충분히 이와 같은 시스템을 구축할 수 있게 되었다. 위 시스템을 통해 생산된 데이터는 불특정 다수에게 수집된다는 장점과 장기간에 걸쳐 계속적으로 수집할 수 있어서 동일인의 재방문

등에 대한 분석이 가능하게 한다. 이를 통해 특정 공간에 대한 선호도를 파악하여 유사패턴 고객들 간의 장소에 대한 추천 서비스를 제공할 수 있다. 기존에 가장 유용하고 널리 사용되는 추천서비스는 바로 사용자 협업필터에 의한 추천이다. A와 B 고객이 유사 고객집단에 속한 고객일 경우 A의 방문 장소 중 B가 방문하지 않은 장소들에 대해 B에게 추천하는 방식이다. 이와 유사한 추천 방식이 바로 콘텐츠 기반 협업 필터링 추천 시스템(박소현, 박영호, 박은영, 임선영, 2018)이다. 이는 A가 방문했던 장소들과 유사한 장소를 추천하는 방법으로 이때 주로 사용하는 기법이 코사인 유사도(Cosine Similarity) 방법이다. 따라서 추천의 성능을 극대화하기 위해선 콘텐츠 간의 유사도 분석의 정확도를 개선하는 것이 중요하고 유사도 분석의 분해능이 높을수록 협업 필터 모델에 의한 추천의 성능도 높아진다고 할 수 있다.

기존의 연구에선 동일한 고객이 동일 매장에 반복적으로 방문했을 때 매장내 체류장소와 체류 시간에 대한 유사도 분석을 진행하였다(김선동, 이재길, 2016). 또한 코사인 유사도 계산 시 체류 시간만을 속성변수로 활용하였고, 체류시간 벡터를 정규화 하여 사용하였다. 그러나 방문에서 체류하는 평균 센서가 많지 않은 점이 유사도 측정 결과에 영향을 주고, 체류시간이라는 단일 속성값으로만 유의미한 결과를 얻기엔 어려웠다.

하여 본 연구에서 기존 연구에서 개별적으로 사용되었던 시간대별 방문자 수와 시간대별 방문자 평균 체류시간, 그리고 체류시간 분산을 이용하여 3가지 입력 속성을 이용하여 지역간, 업종간 유사도 분석을 진행하였다. 이를 통해 기존보다 좀 더 향상된 유사도 분해능과 기존에는 유사한 지역이나 업종으로 분류되었던 곳이 차이가 큰 지역으로 구분될 수 있고 기존에는 유사점이 없는 지역과 업종이 유사한 것으로 판단할 수 있는 새로운 기준과 관점을 찾아볼 것이다.

## 2.2 코사인 유사도(Cosine Similarity) 분석 및 활용 연구

두 변인 간의 유사도를 분석하기 위한 방법으로 코사인 유사도, 자카드 계수, 피어슨 상관계수 등이 존재한다. 본 연구에서는 추천 분야에서 주로 사용하는 코사인 유사도(Cosine Similarity) 분석 방법에 대해 소개한다.

코사인 유사도는 내적 공간의 두 벡터 사이 각에 대한 코사인을 측정한다. 코사인 유사도가 0에 가까울수록 서로 독립적인 경우를 뜻하며, 1에 가까울수록 서로 완전히 같을 경우를 의미한다. 이때 벡터의 크기는 값에 아무런 영향을 미치지 않는다. 데이터의 상대적 크기로 인해 두 벡터 간의 속성을 파악하기 힘들 때 유용하며, 특히 데이터의 패턴(방향)에만 관심이 있는 경우 좋은 성능을 보인다.

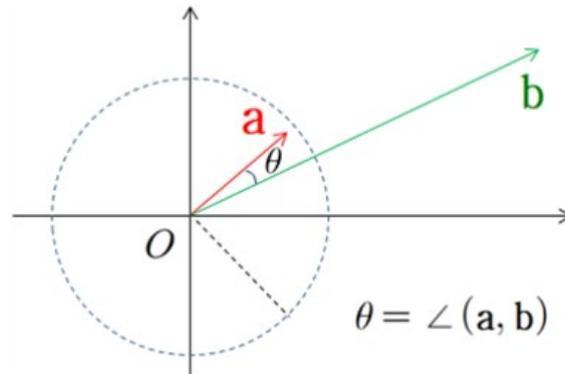
코사인 유사도에서는 데이터의 패턴(방향)에만 관심이 있으므로, 두 데이터(벡터)가 이루는 사잇각  $\theta$ 로 유사도를 측정할 수 있을 것이다. 따라서  $\theta$ 가 작으면 데이터의 유사도가 높고,  $\theta$ 가 크면 데이터의 유사도가 낮다고 판단할 수 있다. 그러나 이 사잇각은 벡터의 내적(inner product)으로부터 정의되므로,  $\theta$ 를 직접 계산하기 보다는 벡터의 내적을 이용하여  $\theta$ 의 코사인 값으로 유사도를 측정한다.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \sqrt{\sum_{i=1}^n (B_i)^2}}$$

두 벡터의 크기는 차이가 크지만, 방향성이 유사하여 그 사잇각이 작고 이를 코사인 값으로 변환하게 되면, 두 벡터 간의 방향성의 유사도를 측정할 수 있다. 벡터간 내적은 다음과 같이 정의된다.

$$a \cdot b = \|a\| \|b\| \cos\theta = a_1 b_1 + a_2 b_2$$

이를 도식화해보면 <그림 1>과 같다. 좌표평면상 벡터로 표현된 두 데이터  $a = (a_1, a_2)$ ,  $b = (b_1, b_2)$ 일 때,



<그림 1> 코사인 유사도 개념

벡터의 내적 계산은 파이썬(Python) 프로그램내에서 '@'연산자 또는 Numpy 라이브러리에 포함되어 있는 dot() 함수를 통해 손쉽게 구할 수 있다는 장점이 있다. 벡터의 크기 또한 norm() 함수를 통해 파이썬에서 계산할 수 있다. 또한, 만약 두 a, b 벡터가 1 차원 row 벡터로 구성되어 있다면, 파이썬 scikit-learn 라이브러리나 scipy 라이브러리에 포함되어 있는 cosine\_similarity(a, b) 함수를 통해 별도의 함수화나 계산과정 없이 바로 이용할 수 있다는 장점이 있다. (<표 1> 참조)

### <표 1> 코사인 유사도에 대한 파이썬 코드 예시

---

#### 코사인 유사도에 대한 파이썬 코드

---

```
import numpy as np
from numpy import dot
from numpy.linalg import norm

def cos_sim(A, B):
    return dot(A, B)/(norm(A)*norm(B))
```

---

본 연구에서는 각 지역별, 업종별 데이터 샘플 수 차이에 의한 크기의 영향을 제외하고 데이터 패턴분석을 통한 유사도 측정이 필요하고, 특히 다차원 분석에 활용할 수 있다는 장점이 있어 코사인 유사도를 주로 사용하였다.

코사인 유사도는 적용이 쉽고, 특히 다차원 벡터에도 적용이 가능하다는 장점이 있어서 다양한 연구분야에서 사용되고 있다. 텍스트 분석(황치곤, 윤창표, 윤대영, 2021)이나 행동 패턴 연구(한이철, 2017; 김경민, 김동윤, 이지형, 2014)에서도 활용되고, 사용자의 위치를 추정(장한메, 김지영, 유기윤, 2017)하거나 매장간 유사도 분석(박소현, 박영호, 박은영, 임선영, 2018)에도 적용되었다. 게임 테스트용 AI 개발을 위해 AI 가 게임 프레임마다 발생시킨 행동데이터와 그에 따른 결과, 그리고 실제 인간이 동일한 게임을 진행하면서 발생시킨 게임내 행동데이터를 다차원 변수로 활용하여, 코사인 유사도 분석을 통해 AI 와 인간의 행동에 대한 연구(허민구, 2021)는 게임테스트용 뿐만 아니라 다양한 분야에서도 활용될 수 있는 연구로, 본 연구에서도 지역별, 업종별 방문자 행동패턴에 대한 유사도 분석 시 다차원 벡터를 입력으로 사용하였다.

또한 기존 연구에서는 매장간 유사도 분석 연구에서는 유사도 분석에 범주형 데이터 1 개만을 입력벡터로 활용하고, 코사인 유사도, 자카드 유사도 및 피어슨

상관계수 분석 모두를 활용하여 유사도를 판단하였다(김소진, 2016). 다만, 본 연구에서 활용된 데이터는 자카드 유사도에 적용하기엔 적합한 형태가 아니어서 이부분에 대한 유사도 분석은 본 연구에서 제외하였고 피어슨 상관계수 분석 역시 두 변인간의 선형적 상관도를 측정하는 분석 방법이라 본 연구의 방법과는 일부 입력 속성에 대한 상관도 분석에는 사용되나 본 연구의 주 목적인 패턴 분석에 맞는 방법은 아닌 관계로 제외하였다. 하여 본 연구에서는 코사인 유사도 모델 하나의 방식으로 진행하되 기존 연구와는 다르게 다차원 입력벡터를 활용하여 이를 개선하고자 한다.

### 2.3 Wi-Fi 로그 데이터 활용에 관한 연구

본 연구에서 활용한 Wi-Fi 로그 데이터는 최근에 많은 실내 공간에서 설치되어 운영되는 무선인터넷 서비스 운영 시 각 개인 단말기가 Wi-Fi AP(무선인터넷 중계기)에 접속한 상태를 기록한 텍스트 메시지이다. GPS 신호를 수신할 수 없는 실내공간에서 기존에는 별도의 비콘(Beacon)이나 RFID 센서 등을 설치하는 형태로 고객방문데이터를 수집하는 형태였는데, 최근 거의 모든 실내공간에서 Wi-Fi 서비스를 도입하고 있고 Wi-Fi 접속 로그 데이터 수집은 별도의 과정 없이 수집할 수 있는 장점이 있어서 많은 매장에서 데이터 수집이 용이하고 그에 따른 응용 연구들이 활성화되고 있다. 매장내부에서의 고객 동선 분석을 통한 재방문 예측(박기연, 2019)이나 고객 군집을 통한 매장내 행동 특성 차이에 관한 연구(김소진, 2016)가 대표적 사례이다.

기존의 재방문에 대한 특징과 재방문 패턴 분석 연구(김소진, 2016; 김선동, 이재길, 2016)에 있어서는 약 1 년에 걸친 두개 매장에서 수집된 추적 데이터를 활용하여 연구를 진행하였다. 7 일 주기로 방문빈도가 강하게 나타나는 것으로

나타났고, 재방문의 행동은 이전 방문의 행동과 일관되는 경우가 많으나 체류 구역만큼은 일관된다고 볼 수 없다는 결론을 내렸다. 해당 연구에서는 특정 매장에 재방문한 고객들에 대한 데이터 분석은 수행하였으나, 특정된 하나의 매장 내에서의 분석이라는 범위의 한계가 존재하였다. 특히 재방문의 행동은 여러 행동 요소에 따라 영향을 받을 수 있어 특정된 매장에서 수집된 데이터 분석을 통하여 재방문 고객의 특성을 일반화하기엔 다소 어려움이 있었다. 즉 기존의 연구결과를 토대로 다른 매장으로 확대 적용하여 활용하는 것에는 무리가 있다. 고객 군집을 통한 행동 특성 연구에서도 선행 연구와 같이 특정 매장 2 곳에서 수집된 데이터를 활용하였다는 데이터의 일반성에는 그 한계성을 벗어나지 못했다.

본 연구에서는 이러한 특정 매장에서 수집된 데이터의 한계성을 탈피하기 위해, 매장 Wi-Fi AP 에서 수집된 단말기 정보가 아니라 개인 단말기에서 수집된 Wi-Fi 접속 로그 데이터를 활용하였다. 하여 기존에 매장이라는 장소적 제약환경에서 벗어나 직접적인 움직임의 주체인 고객이 가지고 있는 개인 스마트폰을 통해 데이터를 수집하였다. 이를 통해 기존의 연구와 동일하게 방문데이터와 체류시간에 대한 정보는 획득할 수 있었으나, 기존 연구와 같이 매장에 실내에서의 동선에 대한 데이터는 매장에 설치된 Wi-Fi AP 에 대한 위치정보가 없는 환경에서는 확보가 어려웠다. 다만, 고객의 다양한 매장 방문에 대한 데이터를 통해 지역과 업종에 대한 데이터를 분석하는 것이 본 연구의 목적임으로 이를 수행하는데 어려움은 없을 것으로 예상된다. 또한 기존에 특정 매장에서만 수집된 데이터의 한정성을 벗어나 일반성을 확보함으로써 좀더 고객 방문에 대한 행태 분석에 다양하게 활용될 수 있을 것이다.

## 제 III 장 연구 방법

### 제1절 데이터 수집 및 전처리

본 연구에 사용되는 데이터는 “프리파이(FreeFi)”라는 앱 서비스를 통해 정보제공 동의한 고객 단말기로부터 수집한 Wi-Fi 접속데이터이다. 일반적인 Wi-Fi 접속데이터는 매장내 설치된 Wi-Fi AP 에서 수집된 단말기 정보를 주로 사용하고 있으나, 본 연구에서는 단말기에서 Wi-Fi 접속에 관한 단말기 로그데이터여서 Wi-Fi AP 와의 접속상태(접속시도/Connected/Time-out/Disconnected)와 같은 정보 뿐만 아니라 rssi, ssid 등 접속하는 AP 와의 신호 세기 등 부가적인 AP 에 관련 데이터도 수집이 가능하다. 또한 단말기에서 일반적으로 Wi-Fi 에 접속하는 방식의 로그데이터여서 특정 매장에 국한된 로그데이터가 아니라 일반적인 다양한 매장의 Wi-Fi AP 에 접속하는 경우 동일한 형태의 데이터를 수집할 수 있는 장점이 있다. 물론 Wi-Fi AP 로부터 수집한 접속로그데이터도 동일한 방법으로 활용 가능하다.

#### 1.1 프리파이(FreeFi) 앱 서비스 소개

프리파이 서비스는 스마트폰에 설치하여야 활용할 수 있는 앱 서비스이다. 사용자가 프리파이 앱을 스마트폰에 설치한 후 Wi-Fi 를 통해 데이터를 사용하면 Wi-Fi 데이터 사용량에 따라 고객에게 일정한 포인트를 리워드로 제공하고, 해당 포인트를 이용하여 문화상품권이나 편의점 상품 교환으로 사용할 수 있는 리워드 서비스이다. 또한 사전 제휴 상점으로 등록된 매장의 경우, 이 앱이 설치된 스마트폰 사용자가 해당 상점에 방문하면 별도의 패스워드 입력과 같은 Wi-Fi 설정 절차 없이 Wi-Fi 가 자동으로 연결되어 사용할 수 있는 서비스이다. 이렇게 고객

고유식별코드가 포함된 개인 단말기로부터 수집된 Wi-Fi 접속 데이터를 고객의 매장 방문에 대한 Data로 활용하고자 하는 것이 본 연구의 취지이다. 다만, 현재는 서비스 업체 사업종료로 인하여 더 이상 운영되고 있지 않다.

## 1.2 Wi-Fi AP 접속 로그 데이터 (AP Log Data)

프리파이 서비스를 통해 2018년 2월 1일부터 2019년 7월 31일 총 18개월간 고객의 Wi-Fi AP 접속 로그데이터(<표 2>, <표 3> 참조)를 6억 1천만여건을 수집하였다. 프리파이 앱 서비스를 통해 고객 단말기 자체에서 전송되는 데이터 수집이어서 별도의 센서 설치 등은 불필요하다. 다만, 고객 개인정보 침해를 최소화하기 위해 단말기에서 제공하는 고유 ID와 핸드폰 맥 어드레스(MAC Address)를 활용하여 암호화하여 개인정보 침해를 최소화하였다. 수집되는 데이터의 기준시간은 그리니치 표준시간대 시간을 사용하여 데이터가 저장되어 있어서 데이터 분석시 한국시간대로 변환과정을 거쳐서 사용하였다. 기본적으로 로그데이터는 몽고DB에서 지원하는 bJSON 형태로 약 300GB 크기로 제공되었으나, 좀더 손쉬운 활용을 위해 데이터 전처리 과정을 진행하면서 이를 .CSV 파일 형태로 변환하여 사용하였다.

아래의 데이터 중 본 연구에 필요한 user\_id/ status/ endedAt/ endedAt\_milis/ shop\_idx 총 5개 속성데이터만 추출하여 분석용 데이터로 활용하였다. 좀더 정확한 고객의 위치데이터를 확보하기 위해 loc\_lat 과 loc\_lng 의 데이터도 활용을 검토하였으나, 대부분의 위치 데이터가 매장의 위치데이터와 동일하여 본 연구에서는 AP 로그 데이터로 수집된 위치데이터 대신 매장정보 DB에 포함되어 있는 매장 위치 Data를 활용하였다. 본 연구에서는 특정 속성데이터만 활용하였으나,

Wi-Fi AP 로그데이터에는 신호의 강도, 보안방식 등 다른 정보도 함께 저장하고 있어 다양한 용도로 활용이 가능할 것이다.

또한 총 수집된 6억 1천만여 건 중 파트너 매장으로 등록된 5만 5천여개의 매장에서 수집된 데이터 3천만여건의 로그데이터 만으로도 충분히 분석이 가능한 수준의 데이터 크기로 판단하여 해당 데이터로만 분석을 진행하였다. 이는 전체 수집데이터의 5% 수준에 불과하나, 15만개의 개인 단말기로부터 수집된 3천만건도 이미 충분히 큰 데이터라고 판단하였다. 또한 별도의 외부데이터를 활용하여 매장정보를 추가하여 분석용 데이터의 크기를 확대할 필요성은 없다고 판단하였다. 우리가 얻고자 하는 데이터는 고객이 매장방문 여부에 대한 데이터임으로 그 기준을 고객 단말기가 매장의 Wi-Fi 신호와 정확히 연결된 시점을 매장방문시점으로 삼고, 연결된 후 바로 다음 연결이 끊어진 로그데이터 입력시까지의 시간차를 체류시간으로 판단하였다.

<표 2> AP 로그 데이터 속성

속성	의미	비고
user_id	User unique id	
ap_id	파트너 매장 AP unique id	
ssid	와이파이 SSID	
status	접속상태 값	0 : disconnected 1 : connected 2 : authentication error 3 : connecting 4 : timeout
rssi	와이파이 신호세기	
security	와이파이 보안방식	

endedAt	Status가 발생한 시간	
endedAt_millis	Status가 발생한 시간의 밀리세컨트 값	
ap_idx	파트너 매장 AP일 경우 해당 AP의 일련번호	
shop_idx	파트너 매장의 AP일 경우 해당 매장 일련번호	
mobile_usage	단말기 부팅이후 현재까지 사용한 모바일 데이터 누적량	
wifi_usage	단말기 부팅이후 현재까지 사용한 와이파이 데이터 누적량	
loc_lat	Status 발생시 수집된 user의 위도	
loc_lng	Status 발생시 수집된 user의 경도	
version	해당 로그가 생성될 때의 앱 버전	

### <표 3> AP 로그데이터 예시

#### 로그 데이터 (Mongo DB)

```

_id: "5a71d9f049c9880a5b7b5100", ssid: "ollehWiFi", loc_lat: 35.8166915, bssid: "06:1d:93:c7:37:1d",
wifi_usage: 6580202599, mobile_usage: 39801704, shop_idx: 0, loc_lng: 128.5321192, ap_idx: 0,
endedAt: "2018-01-31T02:26:26.960+0000", location: [35.8166915, 128.5321192], status: 1, rssi: -43,
security: "OPEN", version: 69, endedAt_millis: 1517365586960, user_id: "599d47d903a3c52298080598",
etc_ap_idx: "f35749f25261c94db341462329eaa5a3", ap_idx: 0, objectId: "5a71d9f049c9880a5b7b5100",
_className: "ap_logs_2018_02", createdAt: "2018-01-31T15:00:00.002+0000", updatedAt: "2018-01-31T15:00:00.002+0000"

```

### 1.3 매장정보 데이터 (Shop Data)

프리파이 서비스 운영 시 미리 제휴한 55,866 개의 파트너 매장들에 대한 정보로써 매장 기본정보와 지역정보, 업종정보, 메뉴, 쿠폰 등 다양한 매장관련 데이터들을 확보할 수 있다. (<표 4> 참조) 다만, 본연구에서는 연구목적에 적합한 매장 기본정보와 지역정보, 업종정보만 한정하여 기 수집된 AP 로그데이터와 결합하여 고객 방문 특성을 살펴본다.

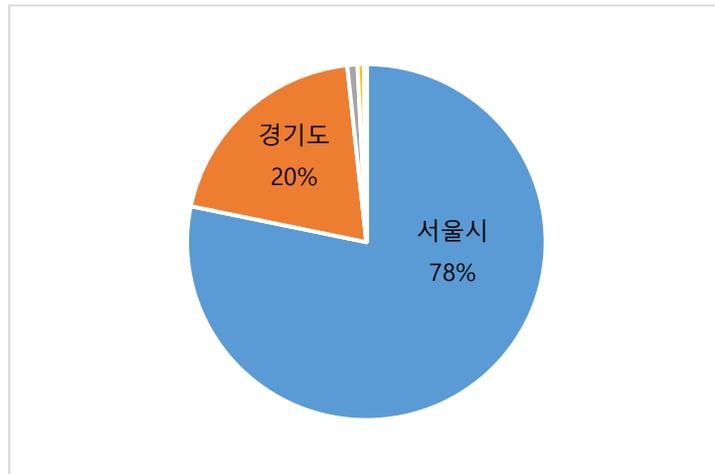
다만, 주소 데이터로부터 지역을 구분하지 않고 사전에 정의된 지역구분 인덱스화 하여 Area 1 카테고리는 ‘특별시, 광역시, 도와 글로벌’ 총 18 개 구분으로, Area 2 카테고리는 ‘시, 군, 구’ 단위로 232 개 인덱스로 구분하였다. 업종별 카테고리도 두단계로 인덱스화 하여 카테고리 1 에서는 ‘기타, 레저, 미용, 술집, 카페/디저트, 패션’ 7 개 업종으로, 카테고리 2 에서는 그 하위 27 개의 세부업종으로 구분하였다.

<표 4> 매장 기본정보

속성	의미	속성	의미
idx	일련번호	Shop_name	매장 이름
categorys_1_idx	카테고리 1 테이블 일련번호	Shop_addr	정규화 주소
categorys_2_idx	카테고리 2 테이블 일련번호	Shop_new_addr	도로명 주소
areas_1_idx	Areas 1 일련번호	Shop2_addr	입력한 주소
areas_2_idx	Areas 2 일련번호	Shop_longitude	매장 경도
Shop_latitude	매장 위도		

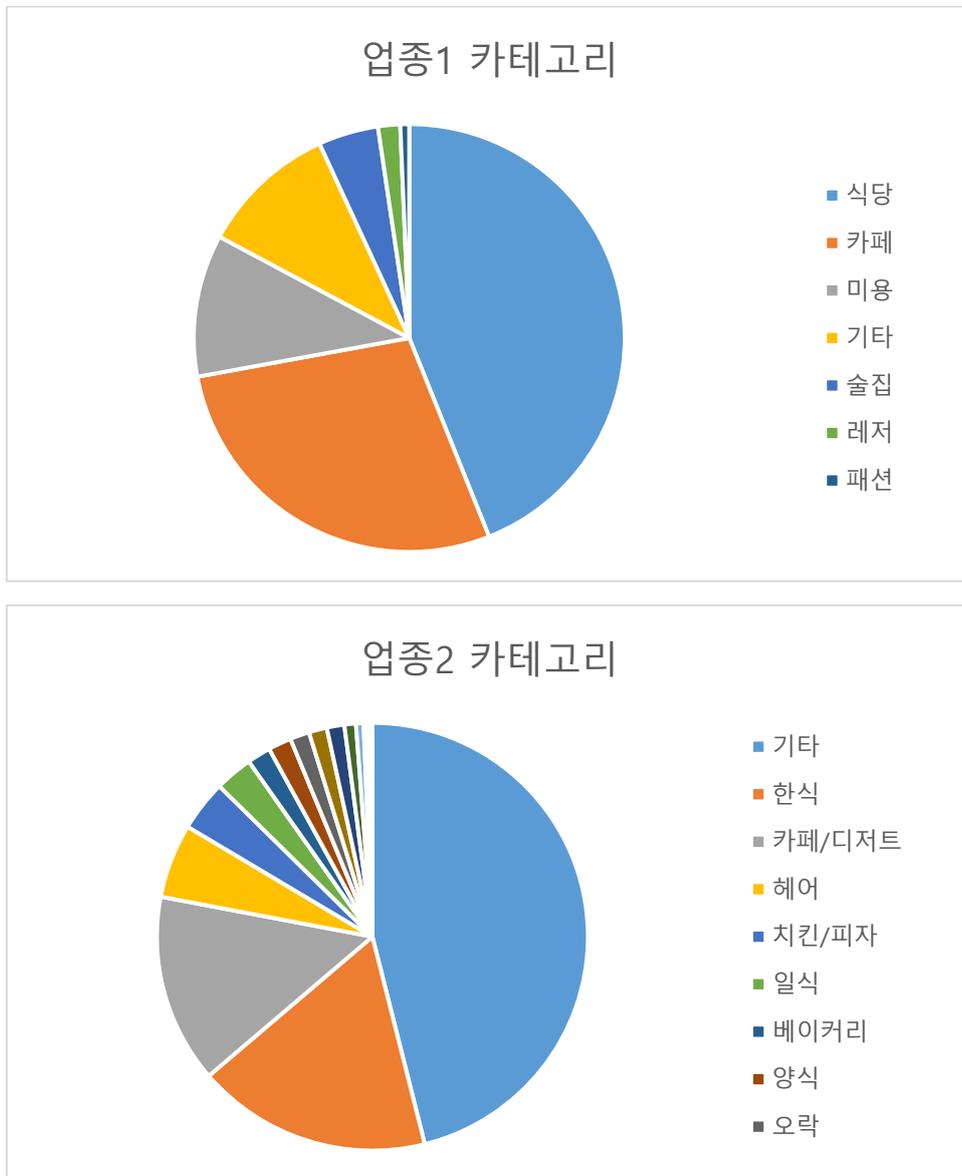
#### 1.4 분석용 마트 데이터 (Mart Data)

위의 1 차 데이터 정제한 Wi-Fi AP 로그데이터와 매장정도 데이터, 지역 범주형 데이터, 업종 범주형 데이터를 결합하여 분석용 Mart 데이터셋을 구성하였다. AP 로그데이터를 중심으로 다른 테이블들을 Left-Join 방식으로 결합시켰으며, NULL 데이터가 포함된 record 는 삭제하거나 적절한 다른 값들로 대체하였다. 이렇게 생성해낸 데이터 마트에 대한 간단히 EDA(Exploratory Data Analysis) 결과, <그림 2>에서와 같이 수집된 데이터의 98%가 서울, 경기지역에 집중되어 있었다. 하여 해당 두 지역의 데이터로만 분석을 진행하였다.



<그림 2> 매장데이터 지역별 분포 현황

업종별 범주형 데이터는 대분류나 소분류 모두 적절한 수준으로 데이터가 분포되어 있는 것으로 확인되어 그대로 활용하였다.(<그림 3> 참조)



<그림 3> 업종 범주형 데이터 분포 현황

일부 이상치 제거 등의 추가 데이터 정제를 거쳐 총 102,784 개의 고유 사용자 식별코드와 총 1,867,790 개의 방문레코드, 25,470 개의 상점정보로 분석용 마트 데이터를 구성하였다. (<그림 4>, <그림 5> 참조)

```
df2.nunique()
```

```

user_id          102784
endedAt         1867790
stay_time       768720
shop_name       25470
shop_addr       21919
shop2_addr      23046
shop_latitude   22833
shop_longitude  22834
a_name_1        2
a_name_2        46
c_name_1        7
c_name_2        17
visit_time     1867790
year            2
month           12
hour            24
dtype: int64

```

<그림 4> 분석용 마트 데이터셋 종합

user_id	endedAt	stay_time	shop_name	shop_addr	shop2_addr	shop_latitude	shop_longitude	a_name_1	a_name_2	c_name_1	c_name_2	visit_time	year	month	hour		
0	5a4fd1b03a3c5671bbe811a	2018-01-31T15:01:51.109Z	1.152900	크리스기제어	서울 중랑구 연북동 461-31	서울시 중랑구 연북동 461-31	37.582093	127.089733	서울시	중랑구	미용	헤어	0001:51.109000+0000	2018-02-01	2018	2	0
1	58e0267b8e0b0dbc3fa65fbd	2018-01-31T15:23:16.188Z	1.291783	복은 손만두	서울 서대문구 남가좌동 268-31	서울시 서대문구 남가좌동 268-31	37.571121	126.913857	서울시	서대문구	식당	기타	0023:16.189000+0000	2018-02-01	2018	2	0
2	59d4acfe03a3c519e5c8c132	2018-01-31T15:01:44.954Z	1.049933	우창만	서울 서초구 반포동 107-28	서울 서초구 반포동 107-28	37.498233	126.996806	서울시	서초구	식당	한식	0001:44.954000+0000	2018-02-01	2018	2	0
3	59e77c3903a3c55d2ea5595d	2018-01-31T15:08:25.227Z	1.190163	카피볼	서울 동대문구 용두동 39-106	서울시 동대문구 용두동 39-106	37.575719	127.040436	서울시	동대문구	카페	커피/디저트	0008:25.227000+0000	2018-02-01	2018	2	0
4	5977421e03a3c506d481d034	2018-01-31T15:22:21.474Z	8.480500	테니스애프로	서울 중구 광희동1가 193-7	서울시 중구 광희동1가 193-7	37.564802	127.006904	서울시	중구	택션	기타	0022:21.474000+0000	2018-02-01	2018	2	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1868065	5afcb56503a3c50bad4a65da	2019-07-31T08:32:46.592Z	1.689433	한방경의인생한의원	서울 강동구 영일동 353-3	서울 강동구 영일동 353-3	37.545207	127.143231	서울시	강동구	기타	기타	17:32:46.328000+0000	2019-07-31	2019	7	17
1868066	5aac63903a3c507e6521ef3	2019-07-31T13:58:42.622Z	1.088383	키미노돼지국밥	서울 중랑구 연북동 392-46	서울 중랑구 연북동 392-46	37.571411	127.083857	서울시	중랑구	식당	한식	22:58:42.622000+0000	2019-07-31	2019	7	22
1868067	5b14e48003a3c514883f86b1	2019-07-31T12:58:10.242Z	1.001950	카피게이트 북동관리공원점	서울 양천구 목동 907-7	서울 양천구 목동 907-7	37.536414	126.879483	서울시	양천구	카페	커피/디저트	21:58:10.242000+0000	2019-07-31	2019	7	21
1868068	5b14e48003a3c514883f86b1	2019-07-31T13:03:10.284Z	4.287017	Lowell hair	서울 광진구 중곡동 161-44	서울시 광진구 중곡동 161-44	37.558389	127.080379	서울시	광진구	미용	기타	22:03:10.284000+0000	2019-07-31	2019	7	22
1868069	5bc2163403a3c50bc37789a	2019-07-31T14:07:06.483Z	3.775550	역도날드 안산고깃국점	경기 안산시 단성구 고깃동 705-5	경기도 안산시 단성구 고깃동 705-5	37.312859	126.828621	경기도	안산시	식당	기타	23:07:06.483000+0000	2019-07-31	2019	7	23

<그림 5> 분석용 마트 데이터셋 예시

## 1.5 최소 체류시간 기준

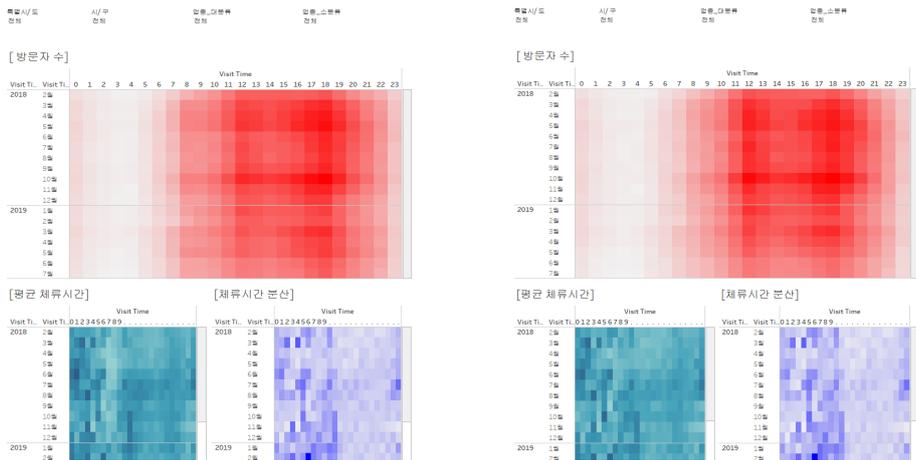
여기서 유의미한 체류시간의 최소단위 기준을 설정하는 것이 중요하다. 너무 짧은 체류시간을 기준으로 하면 데이터의 양은 충분히 크게 확보할 수

있으나, 실제적으로 무의미한 데이터들이 다수 포함될 위험이 있고, 반대로 너무 긴 체류시간을 기준으로 하면 유의미한 고객 데이터가 손실이 될 가능성이 높다. 선행연구에서도 실내 고객 방문 패턴 분석 연구(김선동, 이재길, 2016)에서는 5초로 설정하였고, 매장방문의도 미치는 영향에 관한 연구(박소현, 박영호, 박은영, 임선영, 2018)에서는 관심지역은 10초, 행사 관심고객 구분은 3분이라는 임계 기준을 적용하였다. 기존 선행 연구에서는 고객의 관심에 따른 순간적인 반응도를 측정하기 위해 상대적으로 짧은 체류시간 기준으로 선정하였으나, 본 연구에서는 고객이 매장에 방문한 목적에 따른 행동과 해당 목적을 달성하고 장소를 이탈할 때까지의 적절한 최소한의 체류시간이 필요할 것으로 가정하여 1분과 3분을 기준으로 두 데이터셋을 생성해서 비교해보았다. <표 5>에서와 같이 두 데이터셋의 통계량적 분석에서 차이가 있는 것으로 보이나, 이를 기준으로 두 데이터셋 중 어느 것이 더 적합한 데이터셋이라고 판단하기 어렵고 또한 본 연구에서는 정량적 데이터의 정확성 보다는 데이터 패턴에 대한 분석을 목적으로 하고 있다. 따라서 두 데이터셋의 시간대별 방문자 수나 평균 체류시간의 패턴에서는 크게 차이점이 없고(<그림6> 참조), 패턴 분석에 있어서 집단의 크기가 큰 것이 상대적으로 분석에 유리한 점이 있어 본 연구에서는 1분을 최소 체류시간 기준으로 결정하였다.

<표 5> 최소 체류시간 기준에 따른 데이터셋 비교

임계기준	1 min.	3 min.
count	1,868,070	892,732
mean	13.3	26.0
std	42.68	59.18

25%	1.513	4.726
50%	2.81	9.22
75%	8.57	24.14



<그림 6> 최소 체류시간 기준에 따른 패턴 변화 (1분 vs. 3분)

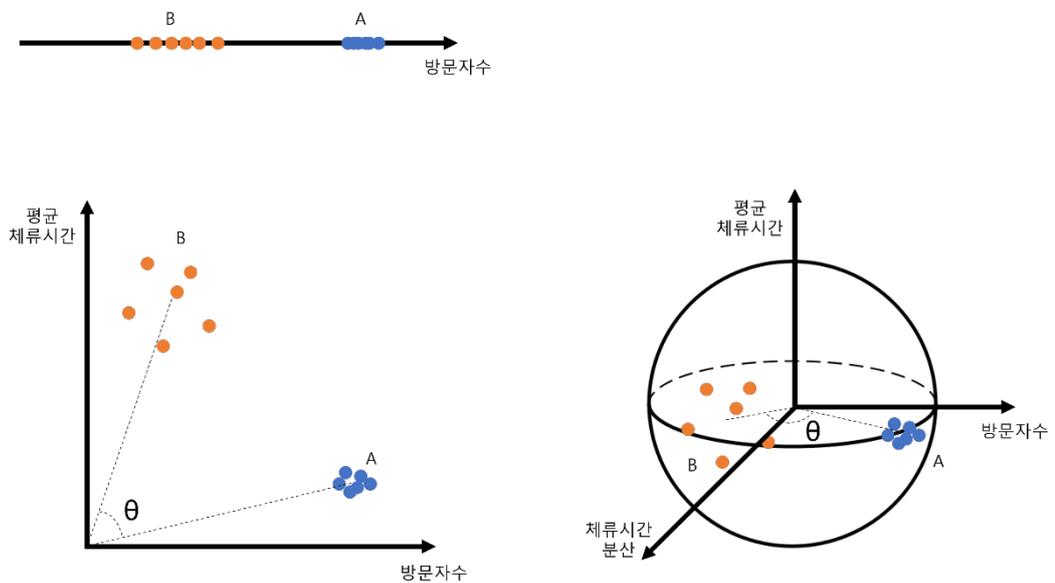
## 제 2 절 데이터 시각화 및 코사인 유사도 모델 설계

본 연구에서는 각 매장의 Wi-Fi 접속 로그 데이터로부터 매장 방문시각과 체류시간 데이터로 변환하여 이를 각 지역별, 업종별, 년, 월, 시간대별로 구분하여 방문자 수와 평균 체류시간, 체류시간 분산을 속성으로 하여 매장 고객이용행태에 대해 분석하였다. 특히 각 데이터의 정량적 수치보단 전체적인 데이터의 패턴에 대한 분석이 본 연구의 주된 목적임으로 데이터 패턴 변화를 확인하기에 적합한 Heatmap 방식의 데이터 시각화 기술을 활용하였고 본 연구에서는 테블로(Tableau) 솔루션 환경에서 구현하였다.

데이터 시각화를 통한 유사 패턴이나 지역의 특이 패턴의 발견에서 그치는 것이 아니라 코사인 유사도 기법을 활용하여 수치화 하여 정량적 분석까지 진행하였다.

기존의 유사도 분석에서도 코사인 유사도 분석기법을 많이 사용해왔었고, 본 연구에서도 동일하게 코사인 유사도 분석을 통해 지역 간의 유사도를 정량화 하여 비교하였다. 다만, 기존의 매장 유사도 분석시에는 ‘방문자수’나 ‘선호도 지수’와 같은 단일 입력변수를 활용하여 측정하였다면, 본 연구에서는 체류시간이라는 새로운 입력 속성과 그에 대한 통계량 데이터를 입력 변수로 활용하여 (방문자수, 평균 체류시간, 체류시간\_분산)과 같이 입력변수를 3 차원 벡터화 하여 유사도 분석이 가능하도록 모델을 설계하여 지역간 유사도를 좀더 세분화하여 검증하였다.

이는 입력변수의 차원을 증가시킴으로써 벡터 간의 상대적 공간 거리 또한 증가시켜 유사 그룹에 대한 분해능(Spatial resolution)을 증가시킨다. ([그림 7] 참조) 단, 차원의 증가로 인해 각 차원 간의 Scale 불균형으로 인해 두 벡터 간의 코사인 계수에 영향을 줄 수 있으므로 반드시 입력변수는 정규화 한 후 사용하였다.



<그림 7> 입력 속성 증가에 따른 코사인 유사도 변화 개념

[기준 모델] A(속성 1), B(속성 1)

[본 연구 모델] A(normalized 속성 1, normalized 속성 2, normalized 속성 3), B(normalized 속성 1, normalized 속성 2, normalized 속성 3)

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \sqrt{\sum_{i=1}^n (B_i)^2}}$$

$$A_{normalization} = \frac{A - A_{min}}{A_{max} - A_{min}}$$

예를 들어, A, B 두매장에 1~3 월까지 각각 다름과 같은 매장 방문고객수라면 A = (1,2,3), B = (10,20,30)인 두 벡터의 코사인 계수 값은 1 로 동일한 패턴을 가진 벡터로 판단할 수 있다. 상기의 두 벡터에 다른 속성의 값을 입력 차원으로 채류시간 데이터를 추가(평균 3, 분산 100)한다고 하면, A = (1,2,3,3,3,3,100,100,100), B=(10,20,30,3,3,3,100,100,100)로 입력 속성을 확장할 수 있을 것이다. 그러나, 이렇게 되면 평균 3 이고 분산이 1 인 동일한 속성이 추가되었기 때문에 A 와 B 의 코사인 계수 값도 이전 분석 결과와 동일하게 1 이 나오길 예상하지만, 계산된 코사인 계수의 값은 0.98180099 로 두 벡터 간에 유사도 특성이 달라졌다는 결과가 나오게 된다. 이러한 문제점을 해결하기 위해서 코사인 유사도 분석의 장점인 벡터의 크기에 상관없이 유사도 분석이 가능하다는 이점을 활용하여 입력 차원에 각 속성별 Normalized 된 값을 사용함으로써 각 입력 속성의 상대적 크기로 인한 코사인 유사도 값의 영향을 최소화하여 이러한 문제점을 해결할 수 있을 것이다. 결국, A = (1/6, 2/6, 3/6, 3/9, 3/9, 3/9, 100/300, 100/300, 100/300), B = (10/60, 20/60, 30/60, 3/9, 3/9, 3/9, 100/300, 100/300, 100/300)의 정규화(Normalized) 된 벡터로 변환되어 두 벡터간 코사인 계수 값은 1 로써 입력 속성이 확장되기 이전의 값과 동일한 상태가 된다. 즉, 우리가 관심있는 것은 결국 데이터의 패턴(방향)에만 관심이 있는 것이라서, 벡터간 비교시에 크기에 관한 영향을 정규화를 통해 제거하고 입력 속성을 증가시킴으로써 유사도에 대한 분해능을 증가시키고자 하는 것이다.

## 제 IV 장 분석 결과

본 연구에서는 지역별 고객 매장 방문 이용 형태 분석을 위해 ‘시간대별 방문자 수’와 ‘시간대별 방문자에 대한 체류시간과 분산’을 활용하였다. 동일한 상품을 판매하며 동일한 시간대에 동일한 방문자 수를 나타나더라도, 해당시간대 방문한 고객의 매장 체류시간이 짧고 빠른 회전을 특성을 보이는 ‘To-go’ 형태인지 그에 비해 상대적으로 긴 매장 체류시간을 가지는 ‘For here’ 유형인지에 따라 고객들의 매장 이용형태가 다르게 나타나기 때문이다. 예를 들면 동일한 커피라는 상품을 구매하는 고객들이 방문하는 매장에서 테이크 아웃(take-out) 서비스를 주로 이용하는 고객이 집중되는 매장과 구매 후 카페에서 마시고 가는 고객 중심의 매장 간에는 고객 매장 이용형태가 차이가 날 것이다. 이를 검증하기 위해 앞서 논의된 분석용 마트 데이터(개인별 매장 방문일시와 체류시간)를 지역별, 업종별, 그리고 년/월/시간대별로 그룹화 하였으며, ‘시간대별 방문자 수’는 각 그룹의 데이터의 개수(Count)로, ‘시간대별 체류시간의 평균과 분산’은 그룹내 체류시간의 평균과 분산으로 설정하여 <그림 8>과 같이 최종 결과 테이블을 생성하였다.

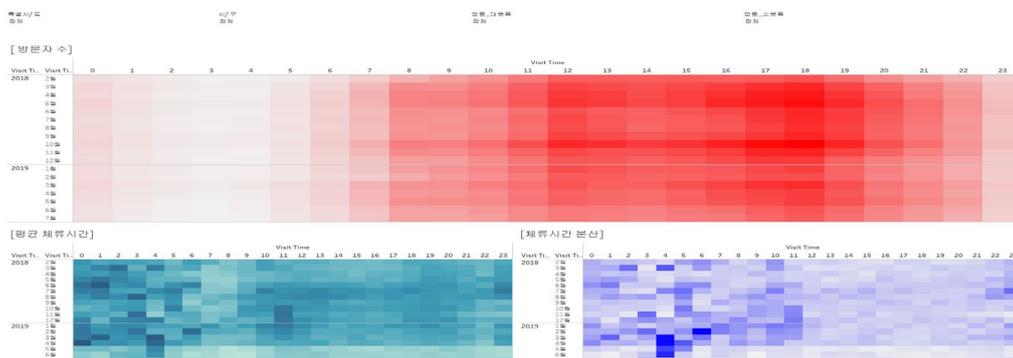
A Name 1	A Name 2	C Name 1	C Name 2	Visit Time (년)	Visit Time (시간)	Visit Time (월)	Stay Time (분산)	Stay Time (카운트)	Stay Time (평균)
서울시	강남구	식당	한식	2018		0	1092.97631544106	5	17.5322566666667
서울시	강남구	식당	한식	2018		1	3947.9429297969	8	34.4586166666667
서울시	강남구	식당	한식	2018		2	30.2960393192659	8	5.47527083333333
서울시	강남구	식당	한식	2018		4	498.862440673472	2	17.2316083333333
서울시	강남구	식당	한식	2018		5	481.844494085555	6	14.5112666666667
서울시	강남구	식당	한식	2018		6	1538.09764560924	11	20.2321606060606
서울시	강남구	식당	한식	2018		7	48194.2603649084	6	97.9102388888889
서울시	강남구	식당	한식	2018		8	7504.58438931152	20	29.2872875
서울시	강남구	식당	한식	2018		9	163.446078481611	20	9.53723833333333
서울시	강남구	식당	한식	2018		10	2093.1703400183	22	25.1033318181818
서울시	강남구	식당	한식	2018		11	6276.94554688441	38	37.4505697368421
서울시	강남구	식당	한식	2018		12	911.849580298272	45	21.5496792592593
서울시	강남구	식당	한식	2018		13	2089.39740207254	30	16.6393066666667
서울시	강남구	식당	한식	2018		14	3025.16120901692	25	27.025342
서울시	강남구	식당	한식	2018		15	3670.22866902718	49	40.5511765306123
서울시	강남구	식당	한식	2018		16	1079.26106654858	37	25.7556617117117
서울시	강남구	식당	한식	2018		17	337.248393205573	32	11.536584375
서울시	강남구	식당	한식	2018		18	1413.72906794855	40	20.6615475
서울시	강남구	식당	한식	2018		19	508.806537596298	33	14.7798313131313
서울시	강남구	식당	한식	2018		20	400.229616657384	31	14.470164516129
서울시	강남구	식당	한식	2018		21	361.759008847988	21	14.6200404761905
서울시	강남구	식당	한식	2018		22	1122.69066883867	16	17.9806666666667
서울시	강남구	식당	한식	2018		23	83.7845462456944	9	8.687

<그림 8> 시간대별 고객 매장 이용 결과 테이블 예시

또한 본 연구의 목적인 데이터 패턴에 대한 차이를 좀더 쉽게 알아볼 수 있게 하기 위하여, 결과데이터의 시각화를 위해 Tableau 상용 소프트웨어를 활용하여 시간대별 방문자 수와 평균 체류시간에 대해 Heatmap방식으로 작성하여 전체적인 데이터 변화의 패턴에 대해 살펴보았다.

## 제1절 지역별, 시간대별 방문 고객 수

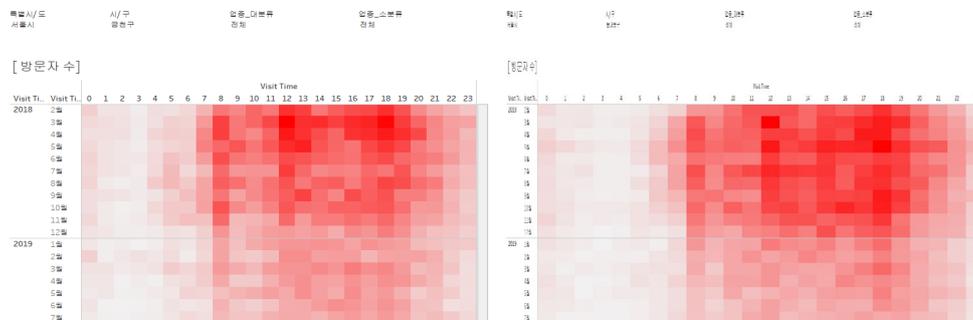
본 연구를 진행함에 있어 각 지역별로 시간대별 고객 매장 이용형태가 다르게 나타날 것이라고 예상하였으나, 전체 분석용 데이터에 대해 종합적으로 살펴보면 <그림 9>에서와 같이 거의 모든 지역에서 오전 8시부터 오후 20시까지 방문객이 주로 발생하였고, 특히 12시와 18시 전후로 집중적으로 발생하는 것을 확인할 수 있다. 대부분의 지역이 위와 같은 방문 데이터 패턴을 보이고 있고 그 주된 이유는 매장 방문의 목적이 주로 ‘음식’과 관련된 업종(술집, 식당, 카페)에 방문한 데이터들이 전체 데이터 중 75% 정도를 차지하여 주로 식사시간대인 12시와 18시 중심으로 매장 방문객이 주로 발생하는 현상이 전 지역 공통으로 나타나는 것으로 예상된다.



<그림 9> 전체 매장 방문 형태

또한 봄(3~5 월), 가을(9~11 월) 시즌에 평소보다 많은 매장 방문자 수가 증가하는 것으로 나타났다. 이러한 패턴은 레저와 패션 업종을 제외한 나머지 업종에 공통적으로 나타난 것으로 이와 같은 시즌성을 감안해 본다면 매장영업전략 수립 등을 진행할 때 분기의 기준점을 봄(3~5 월), 여름(6~8 월), 가을(9~11 월), 겨울(12~2 월)로 설정하여 진행하는 것이 고객 매장 방문 패턴을 고려하여 적합한 마케팅 전략을 수립할 수 있을 것이다.

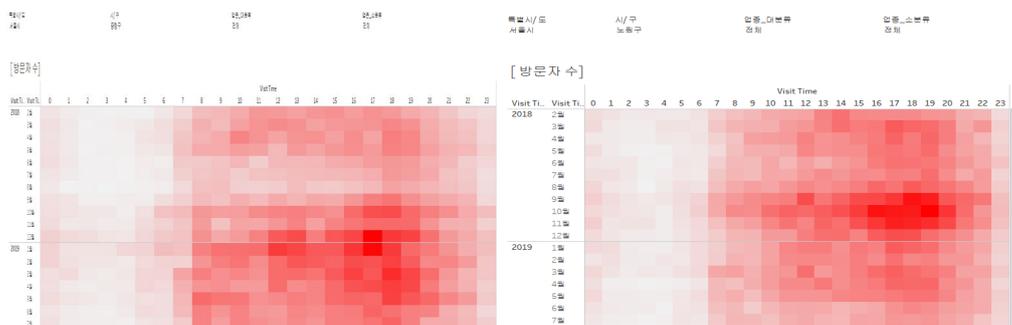
대부분의 지역에선 점심시간(12 시), 저녁시간(18 시)를 중심으로 강한 방문자 패턴이 나타나는 형태이나, 이와는 다르게 좀더 특징적인 패턴을 나타내는 지역도 있었다. 이곳들에 대해 살펴보면 크게 삼시 세끼형/저녁장사형으로 구분할 수 있었다. 12 시, 18 시의 강한 방문 패턴 외에 오전 8 시에도 식당 업종을 중심으로 강한 방문 패턴을 나타내는 지역이 서울시 금천구와 동대문구 등 일부 지역에서 나타났다.(<그림 10> 참조) 이러한 현상이 발생하는 원인에 대해선 외부 자료나 추가적 데이터 결합을 통해 후속 연구가 진행되면 좋을 것이다.



<그림 10> 삼시 세끼형 시간대별 방문자 수 패턴 (서울시 금천구, 동대문구)

이와는 반대로, 저녁시간대(18 시)만 중심으로 강한 방문패턴을 나타내는 저녁장사형 패턴도 중랑구, 노원구, 은평구, 양천구 등 주로 서울 외곽지역에서 나타났다. (<그림 11> 참조) 그리고 이러한 패턴은 미용, 술집 업종에서도 동일하게 나타났다. 이는 서울 외곽지역은 주로 주거 중심 지역이다 보니 상대적으로 회사

등이 밀집되어 있는 상업중심 지역보다 낮시간대의 매장 방문이 낮은 것으로 보인다. 미용과 술집 업종에서도 이와 같은 패턴이 보이는 것 역시 주거와 상당한 관련성이 있는 것으로 판단된다.

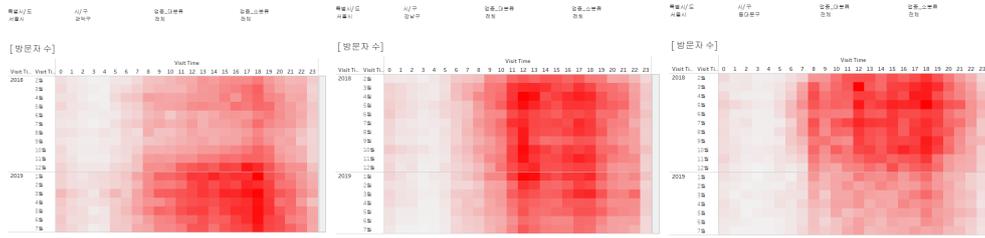


<그림 11> 저녁장사형 시간대별 방문자 수 패턴 (서울시 중랑구, 노원구)

그리고 지역별로 매장 방문 이용형태를 살펴보면, 지역의 상권의 변화를 살펴볼 수가 있었다. <표 6>에서와 같이 2018 년도에는 매장 방문자 수가 많았다가 2019 년도에는 감소하는 지역이 있는가 하면, 반대로 2019 년도에 방문자가 급증하는 지역도 나타났다.

<표 6> 매장 방문자 수 증가도에 따른 지역

매장 방문자 수 증가도	지역
상	관악구, 동작구, 성동구, 영등포구, 중랑구
중	강남구, 강동구, 강북구, 광진구, 노원구, 서초구, 중구
하	강서구, 구로구, 금천구, 도봉구, 동대문구, 마포구, 서대문구, 성북구, 송파구, 양천구, 용산구, 은평구, 종로구



<그림 12> 서울시 관악구 vs. 강남구 vs. 동대문구 방문자수 패턴

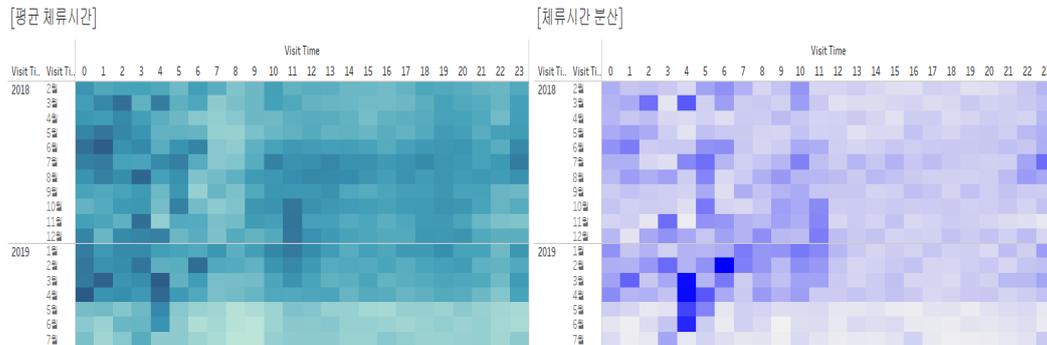
만약 데이터 수집기간을 길게 확대해서 진행할 수 있다면 지역별로 방문자 패턴에 대한 경향성을 파악할 수 있으므로 현재 고객들의 주요 관심지역의 동태적 변화를 살펴봄으로써 지역 활성화 정책 수립이나 상권의 이동 등에 대한 간접적 자료로 활용될 수 있을 것이다.

## 제 2 절 지역별, 시간대별 체류 시간

본 연구에서는 고객 매장 이용형태를 파악하는데 시간대별 방문자수 외에 방문자 체류시간을 고려하였다. 동일한 시간당 방문자 수를 나타내는 매장이라고 하더라도 매장 방문고객의 체류시간이 차이가 있다면 그 매장을 이용하는 목적이나 형태가 다를 것이다. 따라서 본 연구에서는 시간대별 방문자들의 평균 체류시간과 체류시간의 분산을 고객 매장 이용형태 속성으로 활용하였다. 또한 오히려 평균 체류시간보단 체류시간 분산이 오히려 고객의 매장 이용형태에 대한 설명력을 가진다고 볼 수 있어 분산 통계량을 활용하고자 하였다.

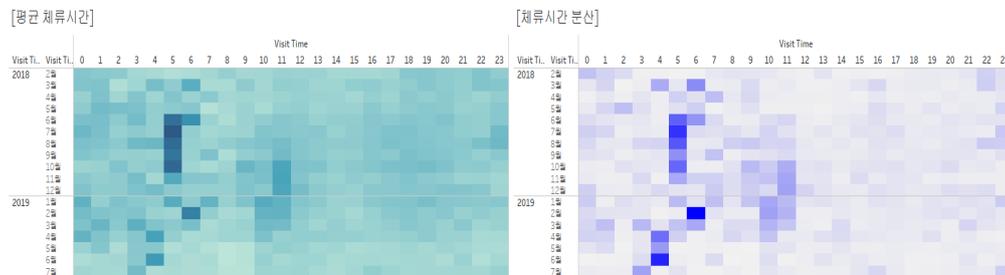
시간대별 평균 체류시간을 살펴보면(<그림 13> 참조) 크게 23시부터 6시까지 야간/심야 시간대의 평균 체류시간이 대체적으로 긴 것으로 나타났다. 이는 평균 방문자 수와 서로 음의 상관관계가 있는 것으로 보이는데, 매장내 혼잡도 감소로

인해 체류시간이 증가된 영향이 있는 것인지 야간/심야 시간대 방문 고객의 특성이 장기체류 특성이 있는 것인지는 추가적인 연구가 필요한 부분이다.



<그림 13> 전체 매장 평균 체류시간 및 분산

‘일상생활 시간대’중에서는 10~ 3 월의 오전 11 시 시간대의 평균 체류시간이 다른 시간대의 평균 체류시간보다 특징적으로 길게 나타났다. 이는 시간대별 특성보다는 ‘식당-한식’업종에서 이러한 현상이 강하게 나타나 전체 데이터에 영향을 미친 것으로 보인다. (<그림 14> 참조)



<그림 14> ‘식당-한식’ 업종 매장 평균 체류시간 및 분산

### 제 3 절 지역 유사도 분석

본 연구에서는 단순히 데이터 패턴에 대해서 시각적 유사성만을 검토하는 것이 아니라 그 유사성에 대한 정도를 수치화 하여 비교하는 것 역시 목적으로 하고 있다. 하여 기존 연구에서도 많이 활용되었던 데이터 패턴간 유사성을 비교하기 좋은 코사인 유사도 계수 분석 방법을 본 연구에서도 활용하였다. 다만, 실내 고객 방문 패턴 분석 연구(김선동, 이재길, 2016)에서는 단순히 시간대별 매장 방문 고객 수라는 1 차원 입력 벡터만을 활용하였다면, 본 연구에서는 매장 방문 고객 수와 더불어 체류시간의 평균과 분산을 활용하여 유사도 분석에 있어서의 분해능을 개선하고자 하였다.

```
# 코사인 유사도 함수
def cos_sim(A, B):
    return dot(A, B)/(norm(A)*norm(B))

# 0 to 1 normalization 데이터 정규화
def normalization(x):
    min_value = min(x)
    max_value = max(x)
    return list(map(lambda x: (x-min_value)/(max_value-min_value), x))

#코사인 유사도 입력 벡터
input_old = result['stay_time', 'count']

input_new_count = normalization(result['stay_time', 'count'])
input_new_mean = normalization(result['stay_time', 'mean'])
input_new_std = normalization(result['stay_time', 'std'])
input_new = input_new_count + input_new_mean + input_new_std

#Result
print(cos_sim(input_old_A, input_old_B))
print(cos_sim(input_new_A, input_new_B))
```

<그림 15> 코사인 유사도 파이썬 구현 예시

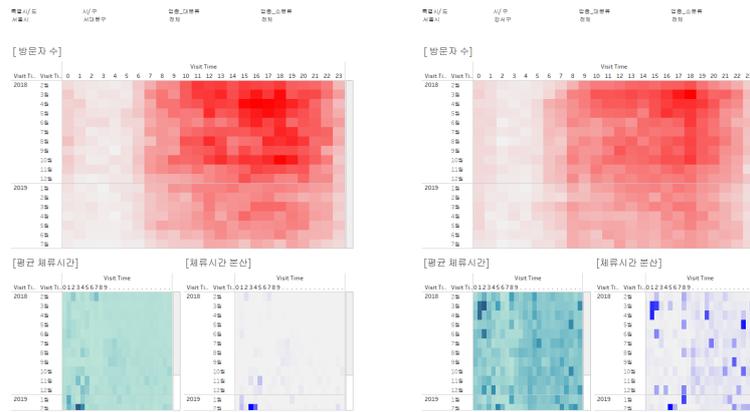
하여 코사인 유사도 함수 입력으로 시간대별 고객 방문 수와 체류시간의 평균과 분산을 결합하여 적용하였고, 파이썬 프로그램에서 이를 손쉽게 구현할 수 있어 이를 통해 각 지역별 유사도 계수를 기존 연구 방식과 비교하여 구할 수 있었다. (<그림 18> 참조)

위와 같은 코사인 유사도 분석을 통해 서울 각 지역별 유사도를 비교해본 결과, 기존의 고객방문자수 단일 속성 유사도와 본 연구 모델인 방문자수, 평균 체류시간, 체류시간 분산 세가지 속성을 통한 코사인 유사도 값의 차이가 가장 큰 지역은 ‘서울시 서대문구’와 ‘서울시 강서구’로서 기존 연구모델에선 0.984 이었던 유사도 값이 본 연구모델에선 0.7783 로써 가장 큰 차이(20.6%)를 나타냈다. (<그림 16> 참조) 또한 두 모델간 유사도 차이가 가장 적은 지역으로 나타난 곳은 ‘서울시 강남구’와 ‘서울시 강동구’로서 기존 모델에선 0.9538, 본 연구모델은 0.9046 으로 약 4.9%정도의 차이를 보이는 것으로 나타났다. (<그림 17> 참조) 이를 시각화 하여 패턴의 차이를 살펴본 결과 서대문구와 강서구는 방문자수에 대해선 비슷한 패턴을 보이고 있으나 평균 체류시간과 체류시간 분산에 대한 데이터 패턴에서의 차이가 큰 것으로 확인되었다. 또한 기존의 연구모델에선 차이가 큰 것으로 보였던 서울시 강남구와 강동구는 오히려 고객 체류시간 속성을 고려하면 유사한 지역으로 나타났다. 심지어 방문자 수 단일 입력 변수로 유사도 분석시에는 서대문구와 강서구의 유사도가 강남구와 강동구의 유사도보다 높은 것으로 나타난 것이다.

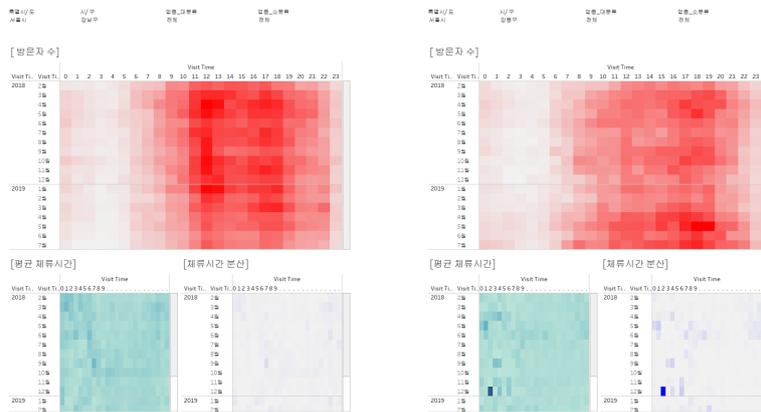
이러한 점들을 살펴보았을 때 기존 연구에서 사용했던 단순 시간대별 방문자 수 단일 입력 변수에 대한 유사도 분석으로는 지역간 매장 방문 고객에 대한 이용형태를 충분히 반영하고 있다고 판단하기 어려움이 있다고 판단된다. 또한 체류시간이라는 변수를 활용함으로써 지역별 유사도 분석에 대한 분해력을 높일 수 있다고 확인할 수 있다.

또한 기존 연구의 코사인 유사도 모델에 있어서 지역간 유사도 계수의 min-Max 편차는 0.166 (0.818~0.984)이고, 본 연구 모델에서의 지역간 유사도 계수의 min-Max

편차는 0.192 (0.713 ~ 0.905)로써 기존 연구모델 대비 15.7%정도의 계수 분해능이 증가됨을 알 수 있었다. 이는 추천 알고리즘에서 유사 아이템 추천 모델에서 정확도에 차이를 줄 수 있는 수준으로 새로운 입력속성을 추가함으로 유사도 측정이 더욱 정밀하게 할 수 있게 됨으로써 협업필터 모델링이나 콘텐츠 기반 추천 알고리즘 등에서 좀더 정확한 성능을 나타낼 수 있을 것이다.



<그림 16> 유사도 편차 최대 발생 지역 비교 (서울시 서대문구 vs. 서울시 강서구)



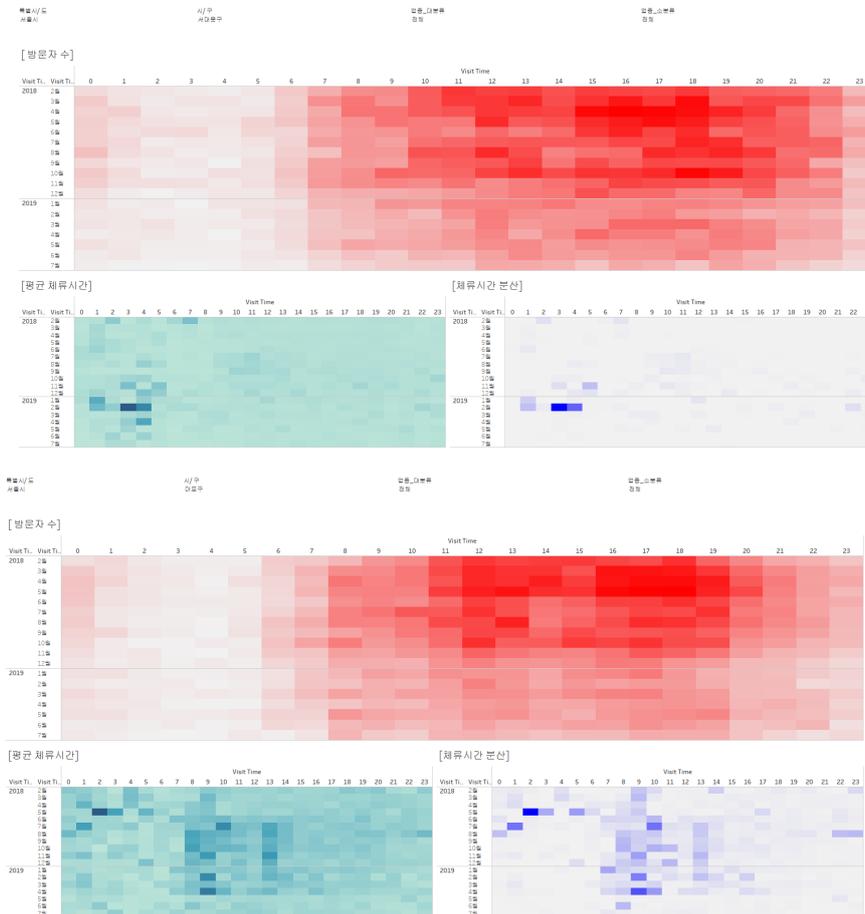
<그림 17> 유사도 편차 최소 발생 지역 비교 (서울시 강남구 vs. 서울시 강동구)

종로구	중구	용산구	성동구	광진구	동대문구	중랑구	성북구	강북구	노원구	서대문구	마포구	양천구	강서구	금천구	영등포구	동작구	관악구	서초구	강남구	송파구	강동구
1.000	0.827	0.863	0.738	0.806	0.851	0.733	0.844	0.808	0.817	0.808	0.884	0.791	0.835	0.841	0.779	0.773	0.758	0.834	0.845	0.842	0.798
0.948	1.000	0.830	0.798	0.796	0.841	0.773	0.845	0.815	0.853	0.840	0.826	0.796	0.798	0.842	0.856	0.837	0.767	0.839	0.904	0.814	0.858
0.961	0.946	1.000	0.755	0.804	0.869	0.739	0.873	0.798	0.834	0.830	0.875	0.812	0.842	0.863	0.788	0.788	0.776	0.823	0.859	0.843	0.809
0.859	0.928	0.879	1.000	0.818	0.760	0.771	0.772	0.831	0.787	0.774	0.757	0.770	0.721	0.761	0.820	0.805	0.784	0.815	0.822	0.750	0.851
0.891	0.929	0.906	0.970	1.000	0.804	0.814	0.817	0.832	0.837	0.765	0.807	0.809	0.822	0.787	0.825	0.837	0.862	0.819	0.833	0.805	0.842
0.967	0.945	0.978	0.873	0.902	1.000	0.727	0.893	0.829	0.863	0.890	0.872	0.833	0.822	0.896	0.799	0.784	0.743	0.835	0.897	0.844	0.849
0.833	0.925	0.851	0.936	0.945	0.852	1.000	0.777	0.772	0.808	0.713	0.756	0.766	0.763	0.719	0.821	0.841	0.826	0.775	0.778	0.742	0.799
0.954	0.944	0.968	0.901	0.934	0.974	0.887	1.000	0.856	0.873	0.891	0.871	0.824	0.837	0.887	0.826	0.824	0.773	0.835	0.890	0.853	0.871
0.880	0.902	0.892	0.951	0.950	0.898	0.895	0.918	1.000	0.837	0.833	0.816	0.804	0.789	0.830	0.827	0.819	0.797	0.855	0.860	0.791	0.898
0.909	0.949	0.927	0.916	0.931	0.939	0.933	0.958	0.916	1.000	0.841	0.833	0.866	0.842	0.837	0.879	0.881	0.842	0.872	0.875	0.852	0.858
0.962	0.941	0.971	0.883	0.918	0.979	0.866	0.978	0.912	0.951	1.000	0.844	0.785	0.778	0.895	0.810	0.773	0.688	0.804	0.908	0.820	0.871
0.976	0.945	0.978	0.881	0.912	0.983	0.845	0.973	0.907	0.930	0.981	1.000	0.818	0.841	0.874	0.783	0.785	0.774	0.826	0.854	0.861	0.816
0.852	0.900	0.894	0.887	0.888	0.907	0.887	0.919	0.884	0.956	0.904	0.887	1.000	0.803	0.785	0.824	0.843	0.797	0.858	0.835	0.799	0.822
0.964	0.950	0.974	0.900	0.933	0.981	0.887	0.980	0.923	0.955	0.984	0.981	0.906	1.000	0.807	0.773	0.792	0.796	0.799	0.808	0.837	0.785
0.961	0.938	0.972	0.879	0.908	0.978	0.844	0.966	0.900	0.917	0.969	0.979	0.865	0.976	1.000	0.790	0.768	0.721	0.811	0.885	0.825	0.851
0.856	0.949	0.871	0.938	0.935	0.867	0.973	0.898	0.901	0.950	0.882	0.862	0.909	0.897	0.854	1.000	0.901	0.837	0.861	0.877	0.778	0.870
0.833	0.933	0.853	0.925	0.919	0.855	0.970	0.886	0.891	0.947	0.866	0.843	0.917	0.881	0.835	0.987	1.000	0.855	0.845	0.852	0.770	0.845
0.818	0.908	0.841	0.954	0.961	0.838	0.978	0.875	0.919	0.915	0.854	0.835	0.880	0.877	0.833	0.966	0.961	1.000	0.819	0.790	0.768	0.795
0.914	0.970	0.935	0.946	0.940	0.934	0.930	0.937	0.927	0.961	0.933	0.930	0.952	0.940	0.916	0.955	0.946	0.929	1.000	0.875	0.821	0.848
0.955	0.976	0.959	0.926	0.948	0.957	0.924	0.960	0.920	0.957	0.959	0.958	0.923	0.965	0.945	0.942	0.928	0.922	0.973	1.000	0.832	0.905
0.971	0.943	0.971	0.882	0.918	0.980	0.867	0.982	0.911	0.948	0.985	0.983	0.899	0.984	0.971	0.878	0.863	0.851	0.929	0.959	1.000	0.804
0.903	0.943	0.911	0.967	0.969	0.914	0.939	0.934	0.969	0.949	0.923	0.917	0.917	0.937	0.909	0.944	0.937	0.952	0.959	0.954	0.925	1.000

<기준 방식 유사도 모델>

<그림 18> 코사인 유사도 모델 기반 지역간 유사도 계수

기존 시간대별 고객방문수로만 비교할 때는 유사한 지역 이었던 ‘서울시 서대문구’와 ‘서울시 마포구’(Similarity=0.981)는 체류시간 속성에서는 서로 다른 이용형태를 보이고 있는 것(Similarity=0.844)으로 나타났다. 특히 마포구의 경우는 <그림 19>에서 나타난 바와 같이 오전 8~10 시까지 긴 체류시간과 체류시간 분산이 크게 나타나 다른 지역과는 차별성이 있는 독특한 패턴이 발견되었다. 이러한 데이터 패턴은 ‘마포구’ 지역의 ‘미용/헤어’ 업종의 오전시간대의 체류시간이 높게 나타나는 경향이 전체 지역 패턴에도 반영된 것으로써 다른 지역에서 나타나지 않은 특징적인 패턴이 발견되었다.



<그림 19> 매장 방문 패턴 비교 (서울시 서대문구 vs. 마포구)

## 제 V 장 결론

### 제1 절 요약 및 시사점

본 연구를 통해 고객 스마트폰 단말기로부터 수집한 Wi-Fi 접속 로그데이터를 통해 지역별, 업종별 매장 방문 고객 특성에 대해 살펴보았다. 우선 Wi-Fi 접속 데이터를 통해 간접적으로 매장 방문 고객들의 특성에 대한 전반적인 패턴을 확인해 볼 수 있었고 이는 우리가 기존에 인식하고 있었던 사실들과 동일한 결과를 보여줌으로써 과거의 설문 조사나 조사원 활용, 별도의 센서 시스템 운영 등의 과대한 비용이 발생하는 조사방식과 동일한 효과를 Wi-Fi 접속 데이터를 통해서도 확인할 수 있음을 검증하였다. 물론 과거의 조사방식이 데이터의 정확성 측면에서는 우수한 점이 있으나, 데이터의 정확성 보다는 데이터 패턴 및 경향성을 중요시하는 본 연구와 같은 분야에 있어서는 손쉽게 확보할 수 있는 Wi-Fi 접속 데이터를 활용하는 것이 사업적으로나 실무적으로도 훨씬 유용한 가치가 있을 것이다. 특히, 마케팅 전략 수립이나 영업 관측 계획과 같은 분야에 있어서는 이러한 중요도는 더욱 중요시 될 수 있다고 예상된다. 또한 본 연구와 같은 방식이라면 과거의 데이터 수집 방식으로는 진행할 수 없었던 대규모 데이터 수집이 필요한 국가 정책 수립이나 지역별 발전 방향 모색 등의 거시경제적 연구에도 활용될 수 있다.

이번 연구에서 중심으로 다루어진 서울 지역에 대한 고객 매장 이용 특성에 대해 살펴보면, 전반적으로 12시, 18시 중심으로 오프라인 매장 방문이 활발히 발생하는 것이 확인되었고, 이는 ‘음식’과 연관성이 높은 매장의 고객 방문이 집중되고 있었다. 또한 봄(3~5월), 가을(9~11월)에 좀 더 많은 매장 방문고객이 발생하는 것으로 파악되었다. 이러한 패턴에 의한다면 주된 영업 관측행사는 시간대적으로는 12시와 18시 중심으로 계절적으로는 봄과 가을에 집중하는 것이 매장 유입 효과를 증대하기엔 좋을 것으로 예상된다. 또한 여름 시즌에 체류시간이

상대적으로 다소 길게 나타나는 경향도 약하게 나타나는 것으로 보여 추가적인 분석을 통해 이를 검증할 수 있다면 여름시즌에는 긴 매장 체류시간을 활용한 매장내 액티비티 행사 등을 통해 매출 증대 효과를 기획해보는 것도 좋은 전략이 될 수 있을 것이다.

매장 내 고객들의 체류시간은 주 활동시간인 8~20시에는 짧고, 저녁~새벽대에는 다소 길게 나타난 것으로 보인다. 다만, 오전식사대에 강한 방문패턴이 보이는 삼시 세끼형 방문패턴이 나타난 금천구와 오전시간대(9~11시)에 체류시간이 긴 고객이 많은 마포구와 같이 독특한 고객 방문 특성이 있는 지역도 존재함이 이번 연구를 통해 발견되었다. 이에 따라 다른 지역과 다소 독특한 매장 이용형태를 보이는 지역에 대해선 다른 지역과는 차별화된 마케팅 전략 수립이 필요한 것으로 보이며, 서울지역 일반적인 고객 특성에 따른 전략은 중앙단위 본사나 관공서 중심으로 수립하고, 차별적 특성이 있는 지역의 고객 이용형태 데이터는 각 지역별 매장 점주 중심으로 이러한 데이터 활용이 가능하다면 좀 더 좋은 매출성으로 이어질 수 있을 것이다.

기존 POI 연구에서 활용되었던 시간대별 방문 고객수를 입력속성으로 활용한 코사인 분석 모델을 개선하여, 평균 체류시간과 체류시간 분산을 추가 입력속성으로 활용한 본 연구의 코사인 분석모델은 지역간 유사도 분석 시 좀더 우수한 분해능을 확보할 수 있어 기존에는 유사 집단으로 구분되었던 지역을 상세히 구분해 낼 수 있게 되었다.

앞서 살펴본 봐야 같이 서울 지역 대부분의 고객 매장 이용 특성이 유사한 형태로 나타나 지역간 유사도 계수 역시 0.7~0.9 수준으로 높은 수준의 유사성이 있음을 확인할 수 있었고, 또한 이번 연구를 통해 각지역의 상대적 거리에 대해서도 확인할 수 있었다. 각 지역 간의 상대적 거리를 확인함으로써 지역 불균형이나 지역 간의 갈등원인을 파악하고 이해하는데 도움이 될 수 있을 것이다.

단일 입력변수를 통해 분석한 코사인 유사도 모델 대비 시간대별 체류시간의 평균값과 분산을 사용하여 세계의 입력변수로 적용해 본 결과 10%이상의 분해능

차이를 파악할 수 있고 이를 통해 지역간 유사도 거리를 좀더 명확히 구분하여 파악할 수 있었다. 따라서 본 연구를 통해 확인된 결과를 유사도 기반 추천 알고리즘인 협업 필터링 알고리즘 등에 적용하게 되면 좀더 우수한 POI 추천결과를 도출할 수 있을 것이다. 또한 기존에 발견하지 못한 독특한 고객 매장 이용 특성을 보이는 지역에 대해선 새로운 추천결과나 영업/마케팅 전략 수립이 가능하게 될 것이다. 이를 확대 적용하게 되면 지역별, 업종별 상권 분석과 같은 거시경제 파악 시에도 본 연구에서 활용한 것과 같이 Wi-Fi AP 접속 데이터에 대한 지속적으로 확보만 가능하다면 별도의 큰 비용 없이 고객들의 매장 이용특성 변화에 대한 동태적 분석도 가능할 것이다.

## 제2 절 연구의 한계 및 향후 계획

본 연구를 통해 Wi-Fi 접속데이터를 통한 새로운 데이터 활용 방법이나 지역간 업종간 유사성 분석 등의 성과를 확인할 수 있었으나, 본 연구에 사용된 데이터셋이 2018년 2월에서 2019년 7월까지 수집된 데이터여서 데이터의 최신성에 있어서 아쉬움이 있다. 특히 2019년 COVID 사태로 인해 전후의 고객 이용형태가 크게 변화되었을 것으로 예상되나 추가적 데이터를 확보할 수 없어서 연구의 범위가 제한적이고 시기성이 다소 늦은 감이 있는 점은 아쉬운 부분이다.

또한 지속적인 데이터 수집을 통해 고객 이용 특성을 동태적으로 분석해 보지 못한 것이 본 연구의 한계성이라고 할 수 있다. 다만, COVID 사태가 점차 이제는 안전화 국면에 접어들고 있고 일상으로의 회복이 진행되고 있는 만큼 기존의 고객 이용형태에 대한 분석이 무의미하다고는 볼 수 없을 것이다. 이번 연구결과를 기준으로 COVID 기간 동안 어떻게 변화하였고, 현재는 COVID이전 대비 어느정도 회복되었는지 확인해볼 수 있는 레퍼런스로 활용도 가능할 것이다. 차후에 본 연구에서 활용되었던 것과 같이 고객의 Wi-Fi 접속 로그데이터를 추가적으로 확보할 수 있다면 매장 방문 고객의 이용 특성의 변화를 동태적으로 확인해 볼 수

있을 것이다.

그리고 본 연구에서 발견한 패턴들에 대한 고객 이용 특성의 원인에 대한 추가적 분석까지 진행되지 못한 것 또한 본연구의 부족한 부분이다. ‘왜 금천구, 동대문구에서는 아침시간대에 집중적인 매장 방문 패턴이 나타나는지?’, ‘서울 외곽 주거중심 지역에서는 왜 저녁식사 시간대에 매장 방문자가 집중되는지?’ 등의 원인에 대한 추가적 분석이 진행된다면 좀더 명확한 고객 이용형태에 대한 이해와 이를 통한 마케팅 전략 수립이 가능할 것이다.

만약 각 지역별 유사도 계수 값을 확인하는 것에서 그치지 않고 고객 방문 이력 데이터를 기반으로 한 지역 군집분석까지 진행하였다면, 지역 추천이나 지역 기반 마케팅을 진행함에 있어서 다양한 인사이트를 도출할 수 있는 결과를 기대해 볼 수도 있었으나, 이는 앞으로 본 연구의 후속 과제로써 진행할 예정이다.

아울러 시간대별, 지역별, 업종별 방문 고객 이용형태를 추가로 확인해 봄으로서 이를 옥외광고 운영에 활용한다면 다른 광고 채널에 비해 발전이 뒤쳐진 옥외광고 채널에 대한 시간대별, 지역별 업종에 대한 타겟 마케팅을 진행할 수 있을 것이고 이에 따른 옥외광고 마케팅 퍼포먼스 개선 정도를 확인해 볼 수 있는 연구로 발전해 나가고자 한다.

## 참고문헌

- 김경민, 김동윤, 이지형 (2014), “트위터를 활용한 감정 기반의 영화 유사도 측정”, *한국지능시스템학회 논문지*, 24(3), 292-297.
- 김다은(2021), 온라인 리뷰에 대한 감성적 언어가 제품 평가에 미치는 영향, 석사학위논문, 성균관대학교.
- 김민재, 이상진 (2014), “코사인 유사도 기반의 인터넷 댓글 상 이상 행위 분석 방법”, *정보보호학회논문지*, 24(2), 335-343.
- 김선동, 이재길(2016), “와이파이 모니터링 기술로 센싱된 고객들의 매장 내부 이동 패턴들을 이용한 재방문 예측”, *한국정보과학회 학술발표논문집*, (), 374-376.
- 김소진(2016), Wi-Fi 로그 데이터를 이용한 매장 방문 고객 특성에 관한 연구, 석사학위논문, 한국과학기술원.
- 김승환, 박진균, 한상용 (2010), “코사인 유사도를 이용한 원자력발전소 운전원 커뮤니케이션 품질 평가 프레임워크”, *한국컴퓨터정보학회논문지*, 15(9), 165-172.
- 김용권(2007), 다차원 데이터를 위한 코사인 유사도를 이용한 검색 공간 축소, 석사학위논문, 한국과학기술원.
- 김혜영, 송지희. (2015), “정보획득 및 구매 단계에서 소비자들의 채널 선택에 영향을 미치는 요인 분석”, *e-비즈니스연구*, 16(3), 3-28.
- 박기연(2019), 웹 로그 분석을 통한 사용자 행동 예측: 온라인 쇼핑 중 구매 결정 지원 기능을 이용한 트래픽을 중심으로, 석사 학위 논문, 연세대학교.
- 박소현, 박영호, 박은영, 임선영(2018), “Point-of-Interest 추천을 위한 매장 간 상관관계 분석 및 선호도 예측 연구”, *한국디지털콘텐츠학회 논문지*, 19(5), 871-880.
- 이세진, 이정교. (2006), “매장 유형에 따른 매장 내 정보탐색 및 구매행동 분석”,

광고학연구, 17(5), 241-253.

이자혜(2018), 온라인, 오프라인의 화장품 구매성향 및 구매형태 연구, 석사학위논문  
숙명여자대학교.

장윤정, 정수아 (2021), “빅데이터를 활용한 외국인 관광객의 국가 그룹별 방문특성  
및 유사성: 트립어드바이저 자료를 중심으로”, *관광연구저널*, 35(3), 95-109.

장한메, 김지영, 유기윤 (2017), “코사인 유사도를 이용한 트윗의 사용자 생성  
공간정보의 중심점 탐지”, *대한공간정보학회 학술대회*, (), 191-192.

장혜원(2021), "정보탐색과 구매 단계별 채널선택의 결정요인." 박사학위논문  
전남대학교.

최성이(2016), 사용자 관심 이슈 분석을 통한 추천시스템 성능 향상 방안, 석사 학위  
논문, 국민대학교.

최영환, 이상용(2006), “오프라인 쇼핑몰에서 개인화된 상품 추천을 위한 사용자의  
이동패턴 분석”, *한국지능시스템학회 논문지*, 16(2), 185-190.

추희정(2022), 온라인 쇼핑몰 고객의 재방문과 재구매 및 이탈 요인 연구, 석사 학위  
논문, 서울과학기술대학교.

한이철 (2017), “코사인 유사도를 이용한 이주패턴 변화의 정량적 측정”, *농촌계획*,  
23(2), 67-74.

황치곤, 윤창표, 윤대열 (2021), “코사인 유사도를 기반의 온톨로지를 이용한 문장  
유사도 분석”, *한국정보통신학회 종합학술대회 논문집*, 25(1), 441-443.

허민구(2021), 코사인 유사도를 이용한 사람과 강화학습 인공지능 게임플레이 분석,  
석사학위논문, 호서대학교

# **Abstract**

A study on regional similarity according to usage patterns of customers visiting offline stores through Wi-Fi AP access log data analysis

Son, Sang Han

Seoul School of Integrated Sciences and Technologies

Advisor: Kim, Bo Young

With the popularization of smartphones, the development of 5G network technology, and the development of big data-related technologies that can process large amounts of data in various forms, it has become possible to collect real-time customers' location and analyze these data by using such techniques. A huge amount of data is collected on the entire purchase decision process from customer entry to final payment in the online shopping mall environment and customer usage patterns are analyzed.

According to data from the National Statistical Office in 2021, the total retail distribution market worth 518 trillion Korean won, otherwise online sales amounted to 187 trillion Korean won. Sales from offline stores still account for a large portion. However, it is harder to collect vast and various types of customer data in an offline store, than in an online store. It is difficult to analyze various customer behaviors based on data as in online stores, and because of this, customer analysis in offline stores is relatively inferior to

customer behavior analysis in online stores.

To solve these problems, studies have been conducted to solve these problems through customer terminal information collected from Wi-Fi APs (Access Points) installed in offline stores. Nowadays, most offline stores have installed and served Wi-Fi to provide various Internet service environments. By collecting Wi-Fi AP access log data, a study was also conducted on the factors that affect the customer's movement in the store for purchase.

However, most studies have controlled data from each single store, and it is difficult to merge data from each store except franchise companies or large distributors. And the form of collected Wi-Fi AP data is different for each manufacturer, so data integration is difficult too. Because of these issues, it is difficult to conduct integrated customer research on multiple stores. Therefore, past studies have been conducted focusing on customer behavior analysis within a single store, and it has been difficult to analyze customer behavior in more macro-regional or commercial district-centered units.

In this paper, rather than customer terminal information collected from in-store Wi-Fi APs in previous studies, Wi-Fi MAC address access log data of visited stores were collected from customers' smartphones and combined with store location and industry data. Differences in customer visit characteristics by time zone, region, and industrial type were analyzed. About 600 million Wi-Fi access data was collected for 18 months, and a total of 1.86 million visit data was refined through a data pre-processing process, and the difference in store usage patterns of customers by time zone, region, and industrial type was analyzed not at each store unit but by region macroscopically. The regional similarity and differentiated customer behavior characteristics of visiting offline stores are shown through the attributes of 'number of customer visitors by time slot(a hour)' and 'store stay time by time slot(each hour)'.

In Seoul, the analysis was conducted by 'gu' unit of each administrative district, and many customer visits occurred in the industrial types for the purpose of 'dining' at 12:00 and 18:00 mainly and it was analyzed that Spring (March-May) and Fall (September-November) showed higher store visit usage patterns than other seasons. In addition, store visit patterns appeared for breakfast (8 o'clock) in some regions and strong store visit patterns were found mainly during dinner hours in the outskirts of Seoul. By utilizing the cosine similarity analysis, which has been widely used in previous studies, this study increased the resolution of the similarity analysis on the usage patterns of customer store visits with the average and variance of the store stay time per time slot in addition to the number of visitors by time slot. If this improved similarity analysis result applied to the region-based recommendation algorithm, it is expected to have a positive effect on improving recommendation performance.

Key words: Wi-Fi, Log Data, Offline store, visit pattern, Time slot, Staying time, Regional Similarity, Seoul

Student Number: 2015406005

## 감사의 글

대학을 졸업한지도 20년이 지나 문득 대학원을 가겠다고 결심을 한지 얼마 되지 않은 것 같은데 어느새 그 마지막에 다가선 느낌입니다. 시작도 그러했지만 잠시 휴학이라는 멈춤이 있었음에도 불구하고 초심을 잃지 않고 끝까지 완주할 수 있었던 것은 항상 주위에서 많은 격려와 도움을 주신 많은 분들이 계신 덕분입니다. 그분들께 지면을 빌어 감사의 인사 전하고자 합니다.

먼저 지난 2년간 부족한 저를 이끌어 주시고 지도해 주셨던 김진호 교수님과 장중호 교수님께 깊은 감사를 드립니다. 그리고 본 논문을 완성하기까지 아낌없는 지도, 편달을 주신 김보영 교수님과 기술적 자문을 해 주신 이래중 교수님께도 감사의 말씀을 드립니다. 바쁘신 와중에도 시간을 할애하여 논문 심사를 진행해주신 신호상 교수님과 임효숙 교수님께도 감사드립니다. 또한 대학원 기간동안 수업 준비와 갑작스러운 코로나 사태로 인한 온라인 수업 전환에도 많은 도움을 주신 김영신 선생님, 김문성 선생님께도 감사 인사 전합니다.

또한 AI 빅데이터 MBA 6기 동기분들이 없었다면 지금의 시간도 존재하지 못했을 것 같습니다. 김현숙 교수님부터 교현이, 상현이까지 6기 동기분들 모두의 지지와 도움 덕분에 잘 마무리를 할 수 있었던 것 같습니다. 특히, 바쁜 직장생활 속에서도 황금 같은 금요일 저녁과 토요일 오후 시간 모두를 ‘방과 후 과정’에 기꺼이 참석해주었던 정경이형, 석인이형, 학진이형, 재필이형, 그리고 두명의 명훈이형, 재홍이형, 그리고 민정이 내외 모두에게 깊은 감사의 마음을 전합니다.

직장생활 중에도 학업에 열중할 수 있도록 배려해주시고 본 논문의 데이터 활용을 허락해주신 (주)디지털투스 김중일 대표님과 업무에 차질이 없도록 늘 배려해 주셨

던 홍승표 이사님, 복인근 이사님, 우동현 이사님께도 이 자리를 빌어 감사의 인사 전합니다. 또한 학교에서 배운 점을 바로 현업에 적용하여 활용할 수 있도록 많은 지원과 아이디어를 주었던 제 생애 최고의 데이터 전문가 집단인 디지투스 멤버 모두에게 많이 늦었지만 이번 기회를 통해 고마운 마음 전합니다.

대학시절에 나중에 논문을 쓰게 되면 서로 이름을 넣어주자고 약속했던 우리 오인방 정훈, 태균, 석민, 병권에게 늦었지만 약속 지켰다고 전하고 싶고, 오히려 고등학교 졸업후에 더욱 친해진 큐피트 멤버들 동욱이, 성재, 진억이, 재경이, 두명의 재욱이, 기운이, 연규, 경훈이, 상일에게도 학업과 직장생활로 지쳤을 때 잠시나마 여유와 웃음을 찾을 수 있게 해주어서 항상 고맙다는 말을 전합니다.

항상 새로운 것에 도전하고 배우는 것에 두려워하지 말라는 가르침과 용기를 주신 아버지와 어머니, 그리고 늘 저를 이해하고 응원해주시는 장인어른, 장모님께 깊은 감사의 마음을 드립니다. 마지막으로 언제나 방향을 잃지 않게 조언해주고 좋은 제안을 해주며 늘 함께 나아가 주고 지지해 주는 아내에게 무한한 사랑과 존경을 담아 고맙다는 말 전합니다. 그리고 하늘나라에서 늘 저를 지켜주고 있는 할아버지와 할머니에게 이 논문을 바칩니다.