

전국 전통시장 시설현황과 리뷰 데이터 분석 보고서

상명대학교 휴먼지능정보공학전공
202210827 김정찬

과제 목표

네이버 지도와 카카오 맵의 리뷰 데이터를 활용하여 고객들이 ‘전통시장’에서 실제로 중요하게 생각하는 시설을 파악하고, 이러한 시설들이 만족도에 미치는 영향을 별점을 통해 분석하는 것을 목표로 한다. 그리고 공공데이터를 이용하여 실제 시설 보유 여부에 따른 각 ‘전통시장’의 만족도 차이를 검증하여, ‘전통시장’의 개선에 있어 시설 투자를 할 때 우선순위를 제시하고자 한다.

데이터 수집 및 전처리 과정

먼저 공공데이터 포털(<https://www.data.go.kr/index.do>)에서 ‘소상공인시장진흥공단_전통시장현황’¹⁾과 ‘소상공인시장진흥공단_전국 전통시장 현황’²⁾을 이용해 1,388개의 전통시장을 매칭해 ‘시장코드’, ‘시장명’, ‘지번주소’, ‘도로명주소’, ‘시도’, ‘시군구’, ‘시장개설주기’, ‘점포수’, ‘개설연도’, ‘취급품목’, ‘아케이드’, ‘엘리베이터/에스컬레이터’, ‘고객지원센터’, ‘스프링쿨러’, ‘화재감지기’, ‘유아놀이방’, ‘종합콜센터’, ‘고객휴게실’, ‘수유센터’, ‘물품보관함’, ‘자전거보관함’, ‘체육시설’, ‘간이도서관’, ‘쇼핑카트’, ‘외국인안내센터’, ‘고객동선통로’, ‘방송센터’, ‘문화교실’, ‘공동물류창고’, ‘시장전용고객주차장’, ‘교육장’, ‘회의실’, ‘자동심장충격기’, ‘공중화장실’, ‘주차장’, ‘위도’, ‘경도’, ‘전화번호’의 컬럼으로 구성된 ‘전통시장_매칭데이터’ 파일을 먼저 만들었다.

매칭하면서 여러 문제가 있었는데 첫 번째로 주소에 결측값이 있거나 중복된 값이 존재하는 문제가 있었다. 두 파일 중 도로명 주소만 있는 데이터와 지번 주소만 있는 데이터가 있었고, 이를 해결하기 위해 두 파일에서 각각 없는 데이터 값을 서로 매꾸는 방식으로 결측값을 처리했다. 두 번째로는 ‘강원도, 강원특별자치도’ 이렇게 같은 지역이라도 다른 표현으로 되어있어 매칭이 안되는 경우가 있었고 ‘address = address.replace('강원특별자치도', '강원도’)’ 형식의 코드로 지역명을 표준화하여 해결했다. 마지막으로는 ‘중앙시장’, ‘신흥시장’과 같은 중복된 시장명이 존재했다. 이를 해결하기 위해서 만약 시장명이 같으면 도로명 주소와 지번 주소 데이터 값을 비교하여 일치한 데이터끼리 매칭하도록 중복을 해결했다. 위와 같은 방법으로 문제들을 해결한 후 괄호, 특수문자, 공백 들과 같은 걸 정리하고 ‘시장명’으로 매칭하여 ‘전통시장_매칭데이터’ 파일을 만들었다.

1) <https://www.data.go.kr/data/15052837/fileData.do>

2) <https://www.data.go.kr/data/15052836/fileData.do>

다음으로 카카오 맵과 네이버 지도에서 각 시장의 ‘별점’과 ‘리뷰’ 데이터를 얻어오기 위해 selenium을 사용하여 동적 크롤링을 진행하였다. 앞서 새로 만든 파일 ‘전통시장_매칭데이터’의 ‘시장명’을 이용해서 검색을 진행하려고 했는데, 몇 번 시도해보니 원하는 장소가 아닌 장소의 데이터를 가져오는 경우가 많았다. 이에 검색 키워드를 수정하여 ‘시군구’ 컬럼과 ‘시장명’ 컬럼을 합쳐서 검색을 진행했다.

```
df = pd.read_csv('전통시장_매칭데이터.csv', encoding='utf-8-sig')
search_keywords = df[['시군구', '시장명']].dropna().apply(lambda x: f'{x["시군구"]} {x["시장명"]}', axis=1).tolist()
```

수정된 검색 키워드를 이용하여 검색한 후 별점과 리뷰가 있는지 확인하고 있는 경우 데이터를 가져오고 없으면 결측값으로 놓는다. 이렇게 갖고 온 데이터들을 정리하여 ‘시장명’, ‘카카오별점’, ‘카카오리뷰’, ‘네이버별점’, ‘네이버리뷰’ 컬럼으로 구성된 ‘review_result_test’ 파일을 만들었다.

이제 완성된 ‘전통시장_매칭데이터’와 ‘review_result_test’ 파일을 이용하여 분석을 진행하려는데, ‘review_result_test’ 파일의 ‘시장명’이 ‘시군구 시장명’으로 이루어져 있고 ‘전통시장_매칭데이터’는 ‘시장명’으로만 이루어져 있어 서로 매칭이 되지 않았다. 이에 중복된 시장이 없으면서 1,388개의 데이터가 잘 유지되도록 ‘시군구 시장명’으로 ‘전통시장_매칭데이터’의 ‘시장명’을 표준화했다.

```
facility_file = '/content/drive/MyDrive/공공빅데이터분석PBL/midterm/전통시장_매칭데이터.csv'
review_file = '/content/drive/MyDrive/공공빅데이터분석PBL/midterm/review_result_test.csv'

df_facility = pd.read_csv(facility_file, encoding='utf-8')
df_review = pd.read_csv(review_file, encoding='utf-8')

df_facility['시장명_표준'] = df_facility['시군구'] + ' ' + df_facility['시장명']
```

이제 두 데이터를 시설 데이터를 기준으로 리뷰 데이터를 left join 하여 결합했다.

네이버 지도와 카카오 맵의 별점의 평균을 내어 ‘종합만족도’ 컬럼으로 사용했고 이에 ‘종합만족도’가 있는 데이터는 총 990개가 나왔다.

```
df_merged['종합만족도'] = df_merged[['카카오별점', '네이버별점']].mean(axis=1)
```

종합만족도가 있는 시장: 990개

또 시설 데이터 유무의 표현이 Y/N으로 되어있는 기존 데이터를 숫자 1/0으로 변환하였다.

```
for col in facility_columns:
    if col in df_merged.columns:
        df_merged[col] = df_merged[col].apply(lambda x: 1 if x == 'Y' else 0)

df_merged[['시장명_y', '종합만족도', '아케이드', '주차장', '공중화장실', '고객유기실', '고객동선통로', '종합글센터', '유애놀이방', '엘리베이터']]
```

	시장명_y	종합만족도	아케이드	주차장	공중화장실	고객유기실	고객동선통로	종합글센터	유애놀이방	엘리베이터
0	가평군 설악눈메골시장	NaN	0	0	0	0	0	0	0	0
1	가평군 가평잣고을시장	NaN	0	0	0	0	0	0	0	0
2	가평군 청평여울시장	NaN	0	0	0	0	0	0	0	0
3	고양시 원당시장	3.935	0	0	0	0	0	0	0	0
4	고양시 능곡시장	NaN	0	0	0	0	0	0	0	0
5	고양시 일산시장	4.305	0	0	0	0	0	0	0	0
6	고양시 덕이동재선1번지	5.000	0	0	0	0	0	0	0	0
7	과천시 과천새서울프라자시장	1.000	0	0	0	0	0	0	0	0
8	과천시 제일쇼핑시장	NaN	0	0	0	0	0	0	0	0

이렇게 최종 전처리 된 파일은 ‘전통시장_최종전처리데이터’이다

분석 방법 및 결과

각 ‘전통시장’에 달린 리뷰 데이터들을 이용하여 고객들이 전통시장에 있는 시설에 대해 실제로 경험하고 느끼는 점을 파악하고자 했다. ‘전통시장_최종전처리데이터’에서 25개의 주요 시설 유무의 대한 정보가 있어, LLM을 활용해 25개의 주요 시설에 대한 포괄적인 키워드 그룹을 정하여 리뷰 데이터에서 해당 키워드들을 얼마나 갖고 있느냐에 따라 시설별 중요도를 정했다.

```
facility_keywords_mapping = {
    '아케이드': ['아케이드', '아케이드형', '실내', '지붕', '덮개', '비가림', '천장', '차양', '햇빛', '그늘', '비오는날', '날씨'],
    '엘리베이터/에스컬레이터': ['엘리베이터', '에스컬레이터', '승강기', '계단', '층', '올라가', '내려가', '위아래', '오르내리', '무거운짐', '계단힘들'],
    '고객지원센터': ['고객지원', '안내센터', '안내소', '정보센터', '도움', '문의', '상담', '서비스센터', '고객센터', '안내데스크', '정보'],
    '스프링쿨러': ['스프링쿨러', '소화', '화재', '안전', '방화', '소화기', '화재예방', '안전시설', '소화설비'],
    '화재감지기': ['화재감지기', '감지기', '화재', '안전', '경보', '화재경보', '연기감지', '알람', '안전장치'],
    '유아놀이방': ['유아놀이방', '놀이방', '아이', '유아', '어린이', '키즈', '놀이', '놀이터', '아기', '육아', '아이들', '어린이시설'],
    '종합콜센터': ['콜센터', '전화', '연락', '상담', '통화', '전화상담', '콜', '전화문의'],
    '고객휴게실': ['휴게실', '쉼개', '쉬는', '았을', '의자', '벤치', '쉼', '쉼터', '앉는곳', '편의시설', '쉴곳', '휴식공간'],
    '수유센터': ['수유센터', '수유', '수유실', '아기', '젖먹이', '모유', '수유공간', '육아', '수유시설'],
    '물품보관함': ['보관함', '물품보관', '보관', '맡기', '사물함', '보관소', '침맡기', '가방보관'],
    '자전거보관함': ['자전거보관', '자전거', '자전거주차', '자전거보관함', '바이크', '자전거거치', '자전거보관소', '자전거주차장'],
    '체육시설': ['체육시설', '운동', '체육', '헬스', '피트니스', '운동시설', '체육관', '운동공간', '헬스장'],
    '간이도서관': ['도서관', '책', '독서', '도서', '서적', '읽기', '도서실', '책읽기', '독서 공간'],
    '쇼핑카트': ['쇼핑카트', '카트', '수레', '끌고', '밀고', '장보기', '쇼핑수레', '장바구니', '카트이용'],
    '외국인안내센터': ['외국인안내', '외국인', '영어', '통역', '번역', '외국어', '국제', '관광안내', '외국인서비스', '영어안내'],
    '고객동선통로': ['통로', '동선', '길', '이동', '다니기', '걷기', '이동통로', '보행', '동선설계', '이동편의', '걸찾기'],
    '방송센터': ['방송', '안내방송', '스피커', '음성', '방송시설', '음향', '알립', '방송안내', '음성안내'],
    '문화교실': ['문화교실', '교실', '강의', '수업', '배우', '문화센터', '강좌', '교육프로그램', '문화활동', '배움'],
    '공동률류창고': ['창고', '률류', '보관', '저장', '률류센터', '창고시설', '보관창고', '률류시설'],
    '시장전용고객주차장': ['주차장', '주차', '차', '자동차', '주차공간', '주차비', '고객주차', '주차시설', '주차면의'],
    '교육장': ['교육장', '교육', '강의실', '세미나', '교육시설', '강의', '교육공간', '교육프로그램'],
    '회의실': ['회의실', '회의', '미팅', '회의공간', '모임', '상담실', '회의기기', '회의시설'],
    '자동심장충격기': ['자동심장충격기', 'AED', '응급', '심장', '응급처치', '의료기기', '구급', '응급의료', '심폐소생'],
    '공중화장실': ['화장실', '변소', 'toilet', '화장실이', '불일', '화장실시설', '위생시설', '편의시설', '화장실깨끗', '화장실더러'],
    '주차장': ['주차장', '주차', '차', '자동차', '주차공간', '주차비', '주차하기', '주차시설', '차량', '주차편의', '주차문제']
}
```

결과는 아래와 같이 나왔다.

===== 리뷰에서 언급된 시설 순위 =====
순위 | 시설명 | 언급횟수 | 주요키워드

1위	주차장	4372회	주차장, 주차, 차
2위	시장전용고객주차장	3726회	주차장, 주차, 차
3위	고객동선통로	290회	통로, 동선, 길
4위	유아놀이방	181회	아이, 어린이, 놀이
5위	문화교실	102회	교실, 수업, 배우
6위	공중화장실	92회	화장실, 화장실이, 볼일
7위	엘리베이터/에스컬레이터	89회	엘리베이터, 층, 올라가
8위	고객휴게실	52회	휴게실, 휴게, 쉬는
9위	아케이드	48회	아케이드, 아케이드형, 실내
10위	종합콜센터	33회	전화, 연락, 상담
11위	쇼핑카트	29회	끌고, 밀고, 장보기
12위	수유센터	28회	수유, 아기
13위	간이도서관	25회	도서관, 책, 도서
14위	외국인안내센터	25회	외국인, 영어, 번역
15위	스프링쿨러	21회	소화, 화재, 안전
16위	고객지원센터	20회	고객지원, 안내센터, 도움
17위	화재감지기	19회	화재, 안전
18위	회의실	18회	모임
19위	방송센터	11회	방송, 음성
20위	공동률류창고	9회	보관, 저장
21위	체육시설	7회	운동
22위	물품보관함	5회	보관, 맡기
23위	자전거보관함	5회	자전거, 바이크
24위	교육장	3회	교육
25위	자동심장충격기	3회	심장

위의 결과를 ‘막대 그래프’와 ‘워드 클라우드’로 표현하면 아래와 같은 결과가 나온다.

중요시 여겨지는 시설

시장전용고객주차장

문화교실

(유아놀이방 서비스 활용도)

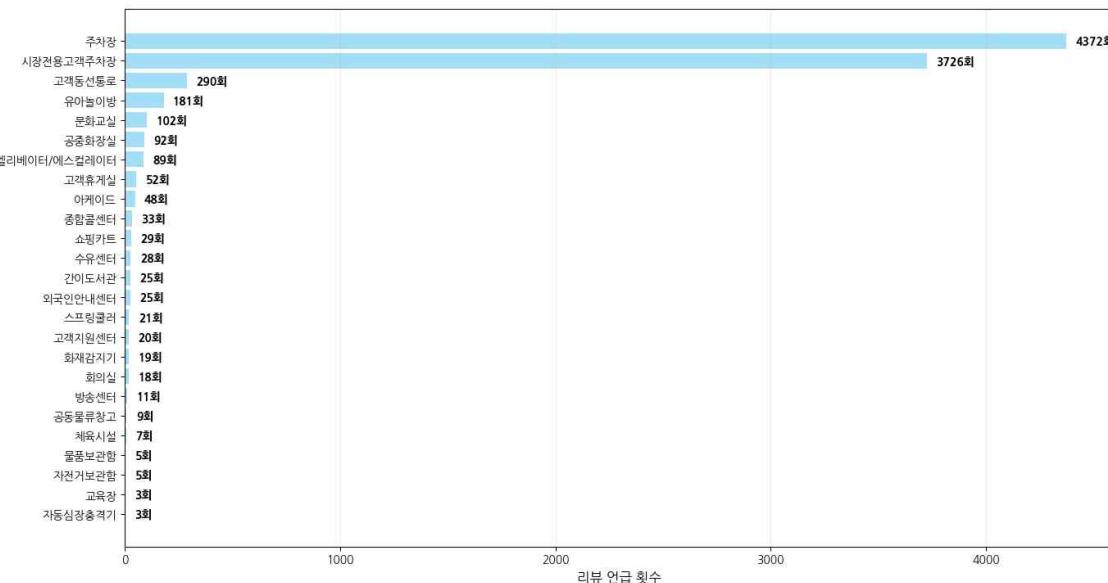
주차장

유아놀이방

고객동선통로

고객동선통로

전통시장 시설별 리뷰 언급 빈도
(고객들이 중요하게 생각하는 시설)



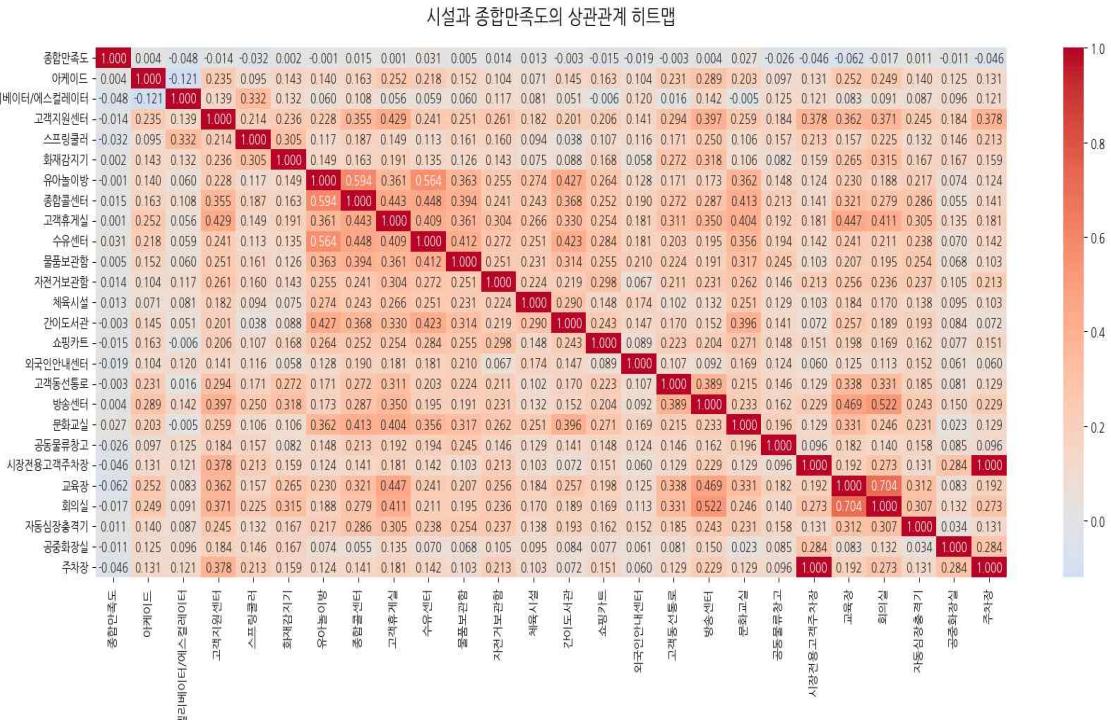
시설 중 ‘주차장’과 ‘시장전용고객주차장’의 언급 빈도가 다른 시설에 비해 월등히 커 위와 같이 조금 치중된 형태를 볼 수 있다.

‘리뷰 데이터’를 통하여 얻은 고객들이 언급 빈도가 높아 중요하게 생각되는 시설 10곳은 ‘아케이드’, ‘주차장’, ‘공중화장실’, ‘고객휴게실’, ‘고객동선통로’, ‘종합콜센터’, ‘유아놀이방’, ‘엘리베이터/에스컬레이터’, ‘문화교실’, ‘시장전용고객주차장’으로 알 수 있었다.

시설명	보유 시장 수	보유비율(%)
공중화장실	948	95.757576
화재감지기	857	86.565657
방송센터	855	86.363636
회의실	837	84.545455
시장전용고객주차장	837	84.545455
주차장	837	84.545455
고객동선통로	808	81.616162
고객지원센터	779	78.686869
교육장	737	74.444444
아케이드	712	71.919192
고객휴게실	551	55.656566
스프링클러	531	53.636364
자동심장충격기	408	41.212121
종합클센터	386	38.989899
자전거보관함	383	38.686869
문화교실	284	28.686869
쇼핑카트	280	28.282828
물품보관함	230	23.232323
수유센터	204	20.606061
유아놀이방	182	18.383838
공동물류창고	176	17.777778
간이도서관	173	17.474747
엘리베이터/에스컬레이터	170	17.171717
체육시설	167	16.868687
외국인안내센터	76	7.676768

다음으로 별점이 있는 시장 990개에 대해서 각 시설 보유 현황과 비율을 알아봤다. ‘공중화장실’, ‘화재감지기’, ‘방송센터’, ‘회의실’, ‘시장전용고객주차장’, ‘주차장’, ‘고객동선통로’, ‘고객지원센터’, ‘교육장’, ‘아케이드’ 이 70% 이상으로 전통시장들이 높은 비율로 보유하고 있는 시설임을 알 수 있다.

다음으로 전통시장의 시설과 종합만족도의 상관관계를 알아 볼 수 있게 히트맵으로 표현해보고 시설 유무에 따른 만족도 차이를 분석해봤다.



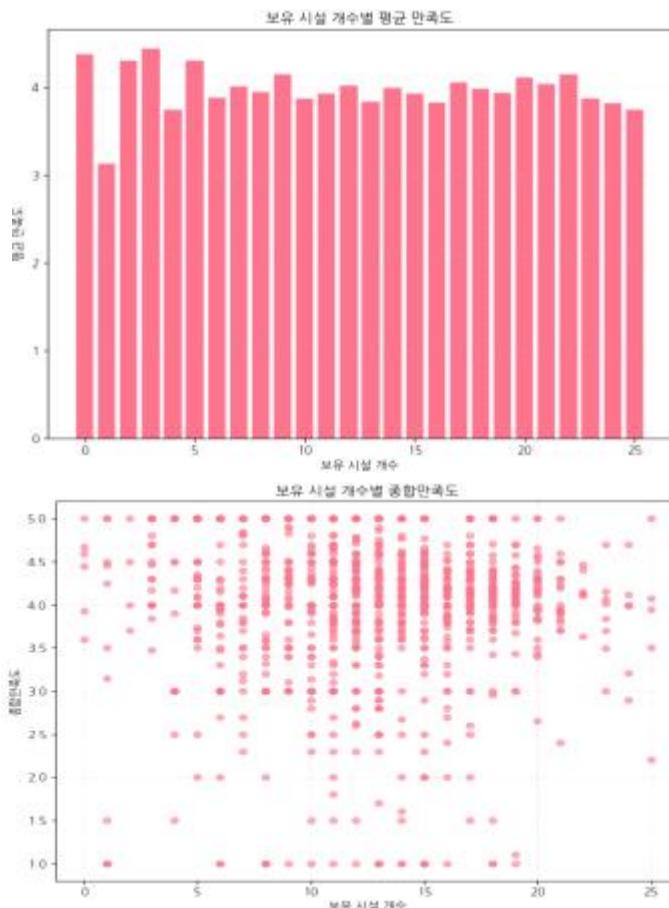
시설	보유유_평균만족도	미보유유_평균만족도	만족도_차이
수유센터	4.020	3.955	0.064
문화교실	4.004	3.954	0.050
체육시설	3.992	3.964	0.028
종합클럽센터	3.985	3.958	0.027
자전거보관함	3.983	3.960	0.023
자동신장충격기	3.980	3.961	0.019
방송센터	3.970	3.959	0.011
물품보관함	3.976	3.966	0.010
화재감지기	3.969	3.965	0.004
고객휴게실	3.970	3.968	0.002
유아놀이방	3.967	3.969	-0.002
고객동선통로	3.968	3.974	-0.006
간이도서관	3.963	3.970	-0.007
쇼핑카트	3.949	3.976	-0.027
고객지원센터	3.963	3.990	-0.028
회의실	3.963	4.002	-0.040
공중화장실	3.967	4.012	-0.046
주차장	3.952	4.058	-0.106
시장전용고객주차장	3.952	4.058	-0.106
엘리베이터/에스컬레이터	3.880	3.987	-0.107
교육장	3.938	4.058	-0.119



분석 결과 큰 상관관계가 압도적으로 큰 변수(시설)은 없었다. 결과를 보면 알 수 있듯 ‘수유센터’, ‘문화교실’, ‘종합 콜센터’, ‘자전거보관함’, ‘체육시설’, ‘자동차장충격기’ 순으로 미세하게 관계가 있음을 알 수 있었다. 하지만 상관관계는 인과관계랑 같이 않으므로 반드시 이 시설들이 종합만족도에 무조건적인 영향을 미치는 것이 아님을 인지하고 결과를 바라봐야 한다.

생각보다 각 시설별 유무에 따른 종합만족도의 관계가 미미하여 추가적으로 총 시설 개수에 따른 만족도를 알아봤다.

총시설 수	평균만족도	세장수
0	4.375000	6
1	3.123182	11
2	4.300000	4
3	4.434474	19
4	3.738333	15
5	4.302500	34
6	3.883333	42
7	4.008125	32
8	3.947830	53
9	4.143878	49
10	3.866455	55
11	3.927432	74
12	4.020119	84
13	3.834722	90
14	3.986319	72
15	3.929041	73
16	3.822100	50
17	4.057705	61
18	3.984537	54
19	3.934595	37
20	4.112419	31
21	4.036316	19
22	4.145833	6
23	3.871250	8
24	3.819167	6
25	3.744000	5



시설이 많을수록 높은 만족도를 받을 것이라는 예상을 했지만 그렇지 않은 결과가 나왔다. 이는 전통시장의 시설 유무가 실제와 다르게 데이터가 입력된 것, 특정 시장에 다른 곳보다 많은 사람들의 평가로 '베스트셀러'처럼 별점이 낮아진 경우 등 실제 상황과는 다른 영향이 있음을 알았다.

결과적으로 '리뷰 데이터'를 통하여 얻은 고객들이 '전통시장'에서 실제로 중요하게 생각하는 시설 10곳은 '아케이드', '주차장', '공중화장실', '고객휴게실', '고객동선통로', '종합콜센터', '유아놀이방', '엘리베이터/에스컬레이터', '문화교실', '시장전용고객주차장'이고, 미미하지만 '전통시장의 시설'이 '종합 만족도(별점)'의 영향을 주는 시설 10곳은 '수유센터', '문화교실', '종합콜센터', '자전거보관함', '체육시설', '자동심장충격기', '물품보관함', '방송센터', '아케이드', '화재감지기'로 비슷하지 않음을 확인할 수 있었다.

결론

과제를 통해 전통시장의 시설과 고객의 만족도(별점/리뷰) 간의 상관관계를 분석했다. 하지만 여기서 몇 가지 이상한 결과를 확인할 수 있었다. 25개의 시설에 대하여 모두 '없음'으로 기록 되어있는 시장에서 높은 만족도를 볼 수 있었고, 일반적으로 예측되는 것과는 다른 상관관계도 볼 수 있었다. 이런 이상한 분석 결과는 공공데이터포털에서 제공되는 전통시장의 시설 정보가 실제와 다르다는 것과 리뷰 수집 과정에서 몇몇 시장은 한 명이 5.0을 줘 만점으로 평가되고 여러 명이 평가하여 한 명의 최하점이 별점을 급격히 낮추는 문제들로 발생했다고 판단한다.

그럼에도 불구하고 현재 수집 및 전처리한 데이터를 바탕으로 한 분석 결과에 따르면, 비록 '리뷰 키워드 분석'과 '시설과 고객만족도 간의 상관관계'는 연관이 없었지만, 리뷰에서 '주차장'과 '시장전용고객주차장'이 압도적으로 언급된 것을 보아 전통시장을 방문할 때 고객이 주차 편의성을 중요한 요소로 생각한다고 볼 수 있다. 하지만 현재 16%의 시장은 주차공간을 보유하고 있지 않는다. 또한 상관관계 수치가 미미한 수준이지만 '수유센터'와 '문화교실'이 전통시장의 시설들 중 그나마 높은 양의 상관관계를 보인다. 이를 통해 가족이 전통시장 방문에 있어 아이와 함께할 수 있는 시설에 대한 수요가 있음을 알 수 있다. 하지만 '수유센터'는 20%, '문화교실'은 28%로 아주 적은 비율의 전통시장만이 해당 시설을 보유하고 있다.

따라서 한정된 예산 및 지원에서 우선적으로 주차장 및 시장전용고객주차장과 수유센터와 문화교실과 같은 가족이 함께할 수 있는 편의시설을 확장하고, 나머지 시설들을 단계적으로 확대하는 것을 통해 전통시장의 만족도를 높일 필요가 있다고 본다.