

# 0428 상권 전처리 현황

사용한 데이터:

## ▼ 서울시 우리마을가게 상권분석서비스

가져온 데이터 및 추출한 column들

### ▼ 생활인구 :

column들이 굉장히 많았다.(532개) → 연령대별, 시간대별, 남성별, 여성별 썩다 나눠진 생활인구수가 나와있었지만, 가장 중요한 “해당 상권의 **총 생활인구수**가 중요해서 아래와 같이 칼럼들을 구성했다.

- 기준\_년\_코드
- 상권\_구분\_코드
- 상권\_구분\_코드\_명
- 상권\_코드
- 상권\_코드\_명
- 총\_생활인구\_수

### ▼ 집객시설 :

집객시설수 외의 세부적인 집객시설의 갯수가 적힌 칼럼들이 있었다. 예를 들면, 초등학교수, 중학교수, 고등학교수, 대학교수, 공항수, 버스터미널수, 정거장수, ... 등이 있었지만, 업종의 분류가 중요하지, 어떤 집객시설이 있는지는 중요하지 않으므로 ‘**집객시설수**’만 추출하였다.

- '기준\_년\_코드'
- '상권\_구분\_코드'
- '상권\_구분\_코드\_명'
- '상권\_코드'
- '상권\_코드\_명'
- '집객시설\_수'

### ▼ 점포 :

점포수 외의 칼럼들인 유사업종점포수, 개업율, 폐업율, 폐업점포수, 프랜차이즈 점포수가 있었지만, null값이 많이있었고, 개폐업율 같은 경우에는 수치가 명확하게

나오지 않아서 제외시켰다. 따라서 **점포수와 서비스업종코드,코드명**만 추출하였다

- 기준\_년\_코드
- 상권\_구분\_코드
- 상권\_구분\_코드\_명
- 상권\_코드
- 상권\_코드\_명
- 서비스\_업종\_코드
- 서비스\_업종\_코드\_명
- 점포\_수

▼ 아파트 :

아파트단지수, 66제곱미터미만의 세대수 , 66제곱미터 세대수, 1억미만 세대수 , 1억세대수, 2억, 4억, 5억, 6억이상 이 있었으나, 아파트의 크기와 가격대에 따른 상권을 분석하기엔 아파트를 제외한 다른 상가의 데이터엔 크기와 가격대에 대한 상세한 정보가 없어서 **평균면적과 평균시가**만 추출하였다.

- 기준\_년\_코드
- 상권\_구분\_코드
- 상권\_구분\_코드\_명
- 상권\_코드
- 상권\_코드\_명
- 아파트\_평균\_면적
- 아파트\_평균\_시가

▼ 추정매출 :

생활인구 데이터처럼 굉장히 많은 칼럼들이 있다. 요일별, 시간대별, 남녀별, 연령대별,,비율 및 건수가 싹 다 칼럼으로 있었지만, 필요한것은 해당상권의 매출금액, 건수이므로 나머지 칼럼들을 제외한 **분기당\_매출\_금액, 분기당\_매출\_건수**를 추출하였다.

- 기준\_년\_코드
- 상권\_구분\_코드
- 상권\_구분\_코드\_명

- 상권\_코드
- 상권\_코드\_명
- 서비스\_업종\_코드
- 서비스\_업종\_코드\_명
- 분기당\_매출\_금액
- 분기당\_매출\_건수

▼ 상권 변화지표 :

다 필요한 내용들 같아서 뺐것은 없다.

- 기준\_년\_코드
- 기준\_분기\_코드
- 상권\_구분\_코드
- 상권\_구분\_코드\_명
- 상권\_코드
- 상권\_코드\_명
- 상권\_변화\_지표
- 상권\_변화\_지표\_명
- 운영\_영업\_개월\_평균
- 폐업\_영업\_개월\_평균
- 서울\_운영\_영업\_개월\_평균
- 서울\_폐업\_영업\_개월\_평균

상권배후지, 상권영역 구분없이 합쳤고, 2017년부터 2022년까지의 데이터를 합쳤다.  
필요한 column들만 추출했다.

---

또한

1. 모든 dataframe의 aggregate될 수 있는 분기별 데이터를 합쳤고 ,
2. 특수문자 포함된 상권코드명 제거하였고
3. 기능상실이 적힌 상권코드명 제거하였고
4. 값이 없는 부분을 key error 전에 제거

5. index를 리셋하였다.

점포와 추정매출은 서비스 업종을 제가 나눴던 분류대로 묶어야 했기 때문에 다시 전처리를 했습니다.

```
음식업_및_주점업 = []
식료품_도소매업=[]
보건업=[]
전문_서비스업 = []
교육_및_사업자원_서비스업=[]
스포츠_및_오락_서비스업=[]
개인_서비스업=[]
기타_서비스업=[]
생활용품_도소매업=[]
가전제품_및_통신기기_도소매업=[]
중고상품_도소매업=[]
기타_도소매업=[]
수리업=[]
숙박업=[]
부동산_및_임대업=[]
제조업=[]
```

위와같이 분류한다음,

```
# 점포에서 없는 서비스업종인 '예술품'-'>'개인_서비스업', '변호사사무소'-'>'전문_서비스업', '당구장'-'>'스포츠_및_오락_서비스업'으로 분류하고 진행했습니다.
for i, j in enumerate(est_sale['서비스업종_코드_명']):
    if(j in ['한식음식점', '호프-간이주점', '치킨전문점', '패스트푸드점', '제과점', '양식음식점', '일식음식점', '중국음식점', '분식전문점', '커피-음료']):
        est_sale['구분'][i]='음식점_및_주점업'
    elif(j in ['반찬가게', '청과상', '수산물판매', '육류판매', '미곡판매', '주류판매', '편의점', '슈퍼마켓']):
        est_sale['구분'][i]='식료품_도소매업'
    elif(j in ['동물병원', '한의원', '치과의원', '일반의원']):
        est_sale['구분'][i]='보건업'
    elif(j in ['세무사사무소', '법무사사무소', '인테리어', '기타법무서비스', '회계사사무소', '통번역서비스', '변리사사무소', '변호사사무소']):
        est_sale['구분'][i]='전문_서비스업'
    elif(j in ['예술학원', '외국어학원', '일반고습학원', '컴퓨터학원', '여행사']):
        est_sale['구분'][i]='교육_및_사업자원_서비스업'
    elif(j in ['골프연습장', '스포츠클럽', '볼링장', '스포츠 강습', 'PC방', '기타오락장', '노래방', '복권방', 'DVD방', '전자게임장', '당구장']):
        est_sale['구분'][i]='스포츠_및_오락_서비스업'
    elif(j in ['세탁소', '피부관리실', '미용실', '사진관', '녹음실', '네일숍', '예술품']):
        est_sale['구분'][i]='개인_서비스업'
    elif(j in ['건축물청소', '독서실']):
        est_sale['구분'][i]='기타_서비스업'
    elif(j in ['일반의류', '운동/경기용품', '유아의류', '한복점', '서적']):
        est_sale['구분'][i]='생활용품_도소매업'
    elif(j in ['핸드폰', '컴퓨터및주변장치판매', '가전제품']):
        est_sale['구분'][i]='가전제품_및_통신기기_도소매업'
    elif(j in ['재생활용 판매점', '중고차판매', '중고가구']):
        est_sale['구분'][i]='중고상품_도소매업'
    elif(j in ['화초', '섬유제품', '완구', '미용재료', '화장품', '문구', '주유소', '전자상거래업', '애완동물']):
        est_sale['구분'][i]='기타_도소매업'
    elif(j in ['모터사이클및부품', '자동차부품', '가전제품수리', '모터사이클수리', '자동차미용', '자동차수리', '통신기기수리', '가방', '신발']):
        est_sale['구분'][i]='수리업'
    elif(j in ['고시원', '여관', '게스트하우스']):
        est_sale['구분'][i]='숙박업'
    elif(j in ['부동산중개업', '가정용품임대', '비디오/서적임대', '의류임대']):
        est_sale['구분'][i]='부동산_및_임대업'
    else:
        est_sale['구분'][i]='제조업'
```

위와같이 구분을 지었습니다.