

빅데이터 2팀_enjoy

상권별 군집분석을 통한

골목상권 활성화 전략 도출

:서울시 요식업을 중심으로

박현기, 이다운, 한규형, 허선영

INDEX

- 분석 배경
- 분석 목적



- 분석 결과
- 활성화 전략 제안
- 한계점

1. 분석 배경 및 목적

2. 데이터분석

3. 분석 결과 및 해석

4. 참고자료

- 데이터 정의
- 데이터 전처리
- 데이터 분석 및 시각화



분석배경 및 목적



분석배경 및
목적



데이터 분석



분석결과 및
해석



참고자료

분석배경

소상공인 현황

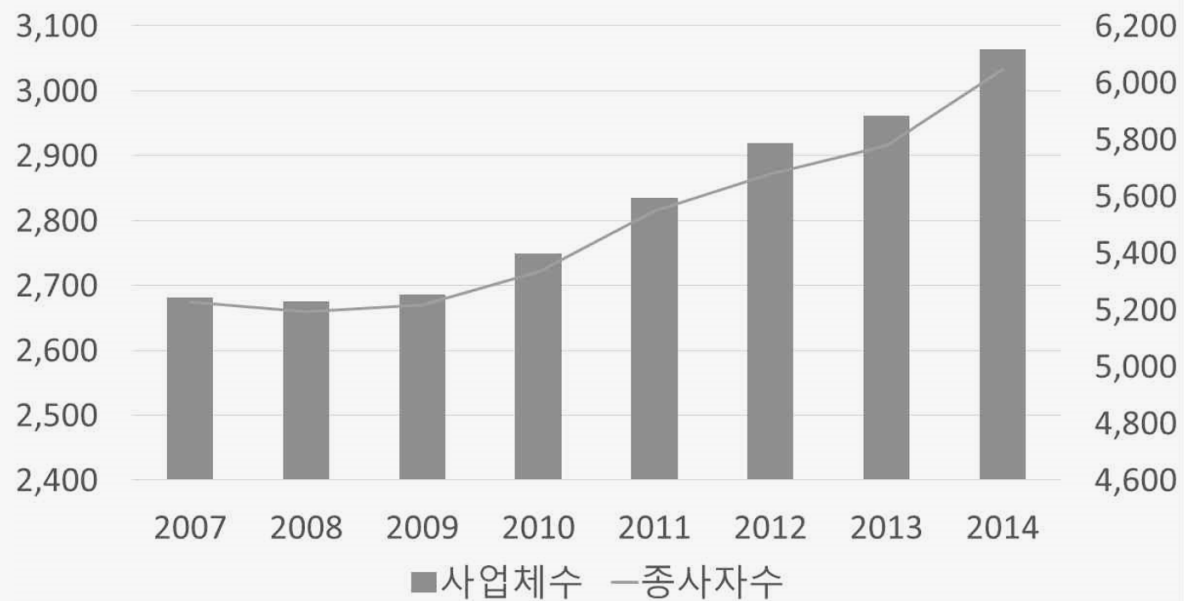
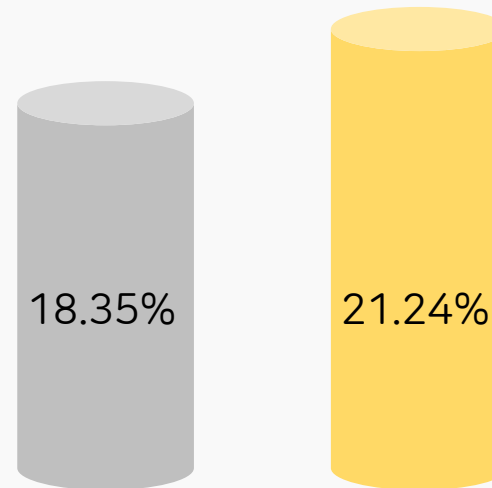


그림 1 소상공인 현황 (천개, 천명)

출처 : 통계청, 전국사업체조사 재편 및 가공

43개 생활밀착형 업종 평균 생존율
(2004년~2014년 기준)



골목상권

발달상권

출처: 서울시(2015)

대형마트 규제에도 전통시장·골목상권 침체 여전

트렌드 모니터

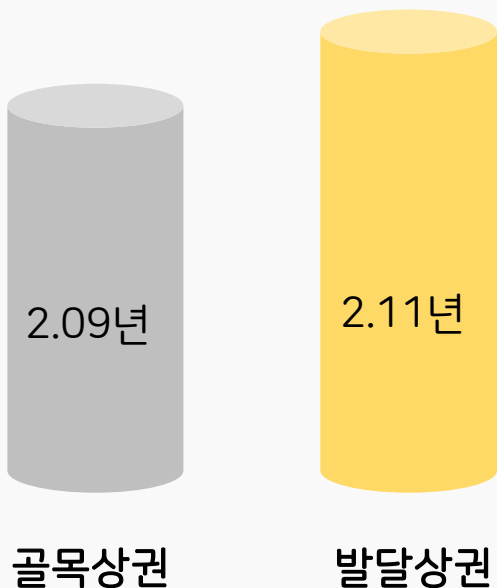
흔들리는 '골목상권', 자영업자들은 <백종원의 골목식당>에서 답을 찾을 수 있을까?

2019-06-20 | 조회수 371 | 공유수 0

가

하

폐업 신고 점포 평균 영업기간



간소비가 위축되면서 개인사업자(자영업자)의 창업 대비 폐업률은 2017년 국세통계 기준 약 76%를 보이며 고용시장 침체, 높은 실업률 같은 사회문제를 야기하고 있다. 특히, 소상공인을 중심으로 이루어진 골목상권의 10년 이상 점포 생존율은 발달상권보다 2.89% 낮게 조사되었고, 폐업점포의 영업기간도 골목상권이 2.09년, 발달상권이 2.11년으로 더 낮은 결과를 보였다(서울시, 2015).

희망하는 창업 분야

(단위 : %, 복수 응답)

분야	비율
1위 외식 및 요리	52.1
2위 온라인 쇼핑	39.0
3위 유통 및 판매	38.8
4위 교육	22.8
5위 지식산업 및 컨설팅	22.4
6위 수공예	21.6
7위 생활 지원	18.1
8위 레저 및 스포츠	17.2
9위 컴퓨터 관련	14.3
10위 대여 및 배달	13.7

자료 : 트렌드모니터

한경BUSINESS

Issue & topic

제 1008호 (20

⊕ 크게 ⊖ 작게 ≡ 프린트 □ 섹션

창업 희망 1순위 '외식업'

전국 성인 1000명 창업 인식 조사... '생계형 창업' 위주 여전

서울시민 39.5%가 외식업 창업 희망

서울에 살고 있는 시민들은 외식업 창업을 가장 선호하는 것으로 나타났다.

서울시가 4일 발표한 창업 수요조사 결과 서울시민이 가장 희망하는 창업분야는 외식업'이라는 답변이 39.5%로 압도적으로 많았다. 이어 유통·서비스업(8.6%), 쇼핑물(6.7%), 패션(5.9%) 등이 뒤를 이었으며, 실버/복지(5.8%)에 대한 관심도 높았다.

분석배경



<발달상권 개념>

발달상권이란 교통이 발달하고 배후 인구와 유동인구가 많은 곳에 형성되는 상권을 의미(정동규, 2017)

서울시의 정의를 기준으로,

‘배후에 넓은 직장 분포, 교통여건양호, 상가와 오피스텔의 높은 밀집, 높은 임대료와 유동인구를 가진 지역’으로 정의



<골목상권 개념>

과거의 모습이 담겨있는 골목길 문화가 남아있으면서 동시에 젊고 창의적인 예술가를 중심으로 만들어지는 상업가(박재홍, 2015)
소규모 개별 브랜드 음식점, 카페 등 영세하고 자본 규모가 작은 상점이 밀집된 지역(정동규, 윤희연, 2017)

서울시의 정의를 기준으로,

‘영세 소상공인이 주로 종사하는 생활밀착형 업종이 밀집한 지역, 일상생활과 밀접한 연관이 있는 상권이 분포한 지역’으로 정의

분석목적



골목상권 침체



지역경제
비활성화



창업실패



일자리문제
심화

골목상권의 침체는 곧 지역경제 비활성화 → 창업 실패 → 일자리 문제 심화로 귀결

∴ 골목상권 활성화를 위한 전략적 제안을 목적으로 함

분석배경 및 목적



분석 배경 및
목적



데이터 분석



분석결과 및
해석



참고자료

데이터 정의

이용 데이터 SET	데이터 기간	공통 변수	사용 변수
상권 추정 매출	2013.01 ~ 2016.12	기준 년월 코드 상권 코드(명)	당월 매출 금액 당월 매출 건수 서비스 업종 코드(명)
상권 추정 유동인구	2014.01 ~ 2016.12		총 유동인구 수
상권 직장인구	2014.01 ~ 2016.12		-
상권 상주인구	2014.01 ~ 2016.12		총 상주인구 수 총 가구 수 아파트 가구 수 비 아파트 가구 수
상권 소득소비	2014.01 ~ 2016.12		월 평균 소득금액
상권 집객시설	2015.01 ~ 2016.12		총 집객시설 수

데이터 전처리

:데이터정제

추정매출금액

	기준_년월_코드	상권_코드	상권_코드_명	서비스_업종_코드	서비스_업종_코드_명	평균_영업_개월수	1년이하_생존율	1~2년_생존율	2~3년_생존율	3~5년_생존율
8263	201301	12621	NaN	CS100007	치킨집	NaN	NaN	NaN	NaN	NaN
8264	201301	12621	NaN	CS100010	호프간이주점	NaN	NaN	NaN	NaN	NaN
8265	201301	12621	NaN	CS200003	예체능학원	NaN	NaN	NaN	NaN	NaN
8266	201301	12621	NaN	CS200004	치과의원	NaN	NaN	NaN	NaN	NaN
8267	201301	12621	NaN	CS200006	일반의원	NaN	NaN	NaN	NaN	NaN

- ✓ 상권 코드 명 변수에 3818개의 결측치 존재 . 다른 테이블에는 존재하지 않는 값이기 때문에 분석에서 제외
- ✓ 평균 영업 개월, 기간 별 생존율 변수들은 절반 이상이 결측치이기 때문에 해당 변수들은 분석에서 제외
- ✓ 데이터 시간적 범위 다른 테이블과 동일하게 설정

데이터 전처리

:데이터 정제

기준_년월_코드	상권_코드	상권_코드_명	총_상주인구_수	총_가구_수	아파트_가구_수	비_아파트_가구_수	
3044	201501	11947	당산로44길	194	84.0	84	0.0
3045	201501	11948	동일로242길	1789	638.0	124	514.0
3046	201501	11949	동일로_A	748	288.0	173	115.0
3047	201501	11950	동일로_B	190	70.0	0	70.0

기준_년월_코드	상권_코드	상권_코드_명	집객시설_수	
0	201501	11947	당산로44길	10
1	201501	11948	동일로242길	11
2	201501	11949	동일로_A	6
3	201501	11950	동일로_B	22

	기준_년 월_코드	상권_ 코드	상권_ 코드_명	총_직장_ 인구_수	남성_직장_ _인구_수	여성_직장_ _인구_수	연령대_10_직 장_인구_수	연령대_20_직 장_인구_수	연령대_30_직 장_인구_수	연령대_40_직 장_인구_수
12096	201501	11947	당산로 44길	375	148	227	0	30	121	121
12097	201501	11948	동일로 242길	2189	1704	485	1	53	165	314
12098	201501	11949	동일로 _A	272	114	158	0	20	62	71
12099	201501	11950	동일로 _B	485	204	281	0	53	133	129

3048

	기준_년 월_코드	상권 코드	상권_코드_명	총_유동 인구_수	남성_유동 인구_수	여성_유동 인구_수	연령대_10_유 동인구_수	연령대_20_유 동인구_수	연령대_30_유 동인구_수	연령대_40_유 동인구_수
12	201501	12169	웅마산로 115길_B	5165	2609	2556	175	572	1044	1247
13	201502	12169	웅마산로 115길_B	4602	2333	2269	159	480	952	1114
14	201503	12169	웅마산로 115길_B	5913	2987	2926	208	593	1210	1385
15	201504	12169	웅마산로 115길_B	5709	2885	2824	197	574	1200	1334
16	201505	12169	웅마산로 115길_B	5909	2951	2958	239	620	1245	1368

기준_년월_코드	상권_코드	상권_코드_명	월_평균_소득_금액	소득_구간_코드	
11952	201501	11947	당산로44길	460	6
11953	201501	11948	동일로242길	301	3
11954	201501	11949	동일로_A	346	4
11955	201501	11950	동일로_B	235	2
11956	201501	11953	도봉로_A	422	5

- ✓ 각 테이블에서 필요한 변수 추출
- ✓ 다른 테이블과 기준 년월 코드 동일하게 설정
- ✓ 상권 코드 명이 존재하지 않는 행 분석에서 제외

데이터 전처리

:데이터 병합및 결측치 처리

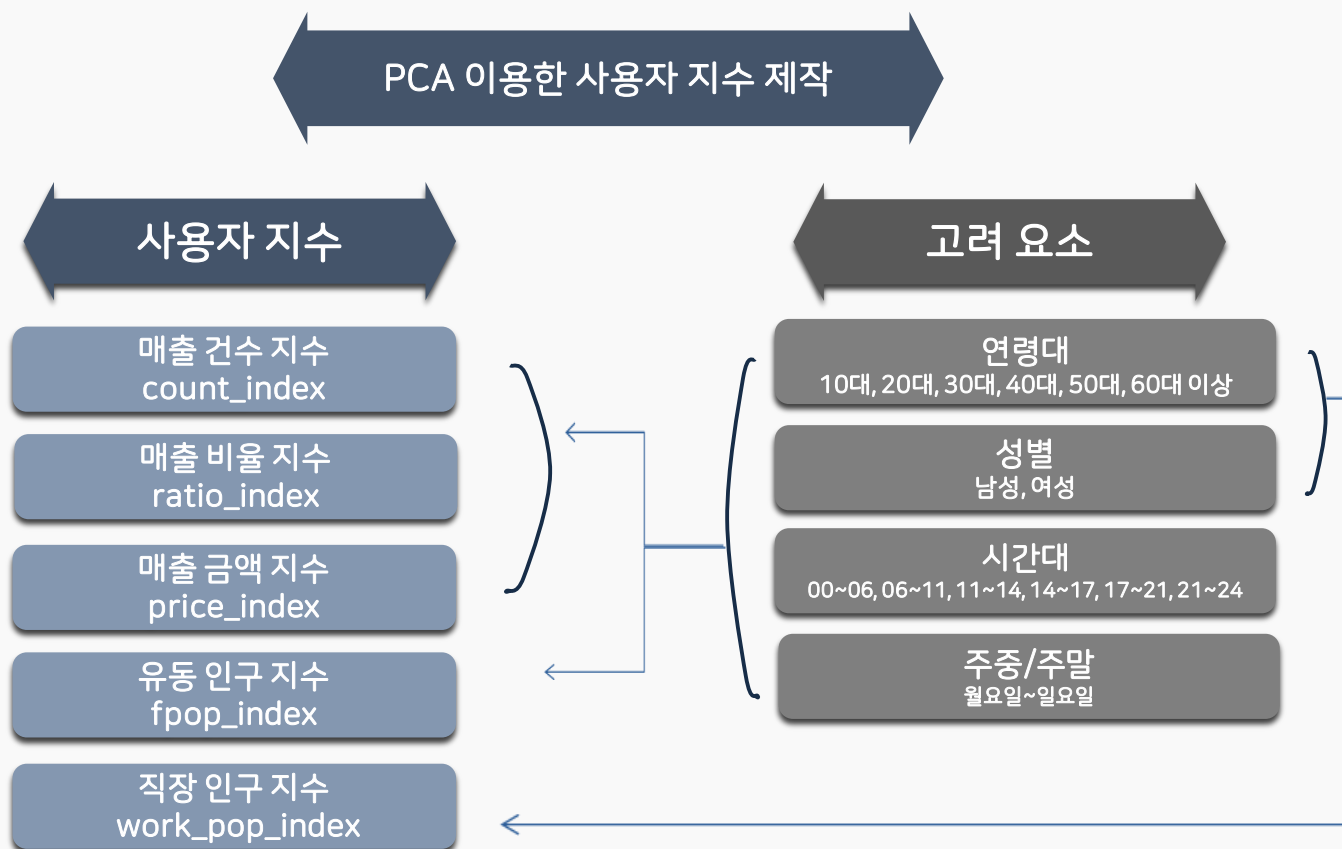
	기준_년 월_코드	상권_ 코드	상권_ 코드_명	서비스_업 종_코드	서비스_업종 _코드_명	당월_매출 _금액	당월_매 출_건수	주중_매 출_비율	주말_매 출_비율	월요일_매 출_비율	화요일_매 출_비율	수요일_매 출_비율
0	201501	11947	당산로 44길	CS100001	한식음식점	96722641	4527	68.7	31.3	16.6	11.8	12.9
1	201501	11947	당산로 44길	CS100004	양식집	14242932	910	77.4	22.6	17.1	13.2	12.7
2	201501	11947	당산로 44길	CS100005	분식집	5084735	653	84.4	15.6	13.2	12.4	18.2
3	201501	11947	당산로 44길	CS100006	패스트푸드 점	18820989	2098	69.6	30.4	11.7	13.2	13.7
4	201501	11947	당산로 44길	CS100008	제과점	124446237	11102	73.6	26.4	13.6	12.6	14.2

- ✓ 상권추정매출 테이블을 기준으로 left merge
- ✓ 병합 과정에서 생긴 결측치를 포함한 데이터는 분석에서 제외

- ✓ 업종이 요식업인 데이터만 사용
- ✓ 최종 데이터 형태 : 171388 rows, 115 columns

데이터 전처리

:차원축소-지수 생성



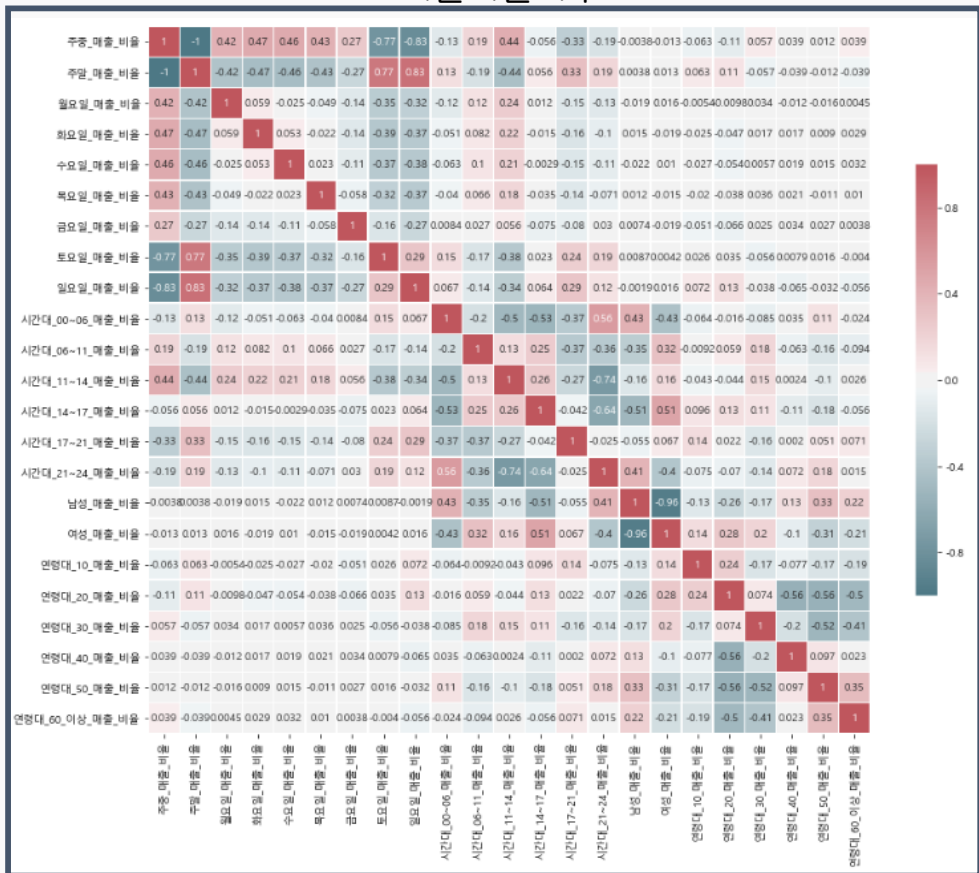
종합지수= $w_1*y_1+w_2*y_2+w_3*y_3+\dots$
(w_i : 주성분별 가중치, y_i : 제 i 번째 주성분 계수)

- ✓ 주성분 분석을 통해 비슷한 특성을 가진 다양한 요소들을 한 번에 고려할 수 있는 지수 생성
- ✓ 여러 변수가 갖는 공통 성분 중 변동성을 가장 잘 설명하는 주성분 추출
- ✓ 각 주성분이 갖는 설명력을 가중치로 곱하여 환산

데이터 전처리

: 차원 축소 - 지수 생성과정

< 매출비율지수 >



```
# StandardScaler (평균을 0 으로 만들어 줌)
scaler = StandardScaler()
scaled_ratio = scaler.fit_transform(market[ratio_columns])

# 주성분의 개수 선택 방법 : 누적 분산 기준 95%
pca = PCA(n_components=14)
pca.fit(scaled_ratio)
```

```
# 누적 고유치 비율
pca_ratio = list(pca.explained_variance_ratio_)
print("1) pca성분 별 비율")
print(pca_ratio)
print("")
print("2) pca성분 별 비율의합 : ", sum(pca_ratio))
print("")
# 가중치 산정
pca_weight = pca_ratio / sum(pca_ratio)
print("3) 주성분 별 가중치 산정 : ")
print(pca_weight)
```

```
ypca = pca.fit_transform(scaled_ratio)
```

```
# 새로운 지표 생성
# 주성분 * 가중치
market['ratio_index'] = (ypca * pca_weight).sum(axis=1)
```

1) pca성분 별 비율
[0.21200749650390147, 0.1666591095728709, 0.09215837147686179, 0.06126380964120195, 0.05418833108461497, 0.050510870670319806, 0.04728065299915654, 0.04550784304316104, 0.0427776943513367, 0.04104601268574037, 0.0395583136811869, 0.030578940059030425, 0.02811465328436042, 0.02624085315118979]

2) pca성분 별 비율의합 : 0.93789302728873

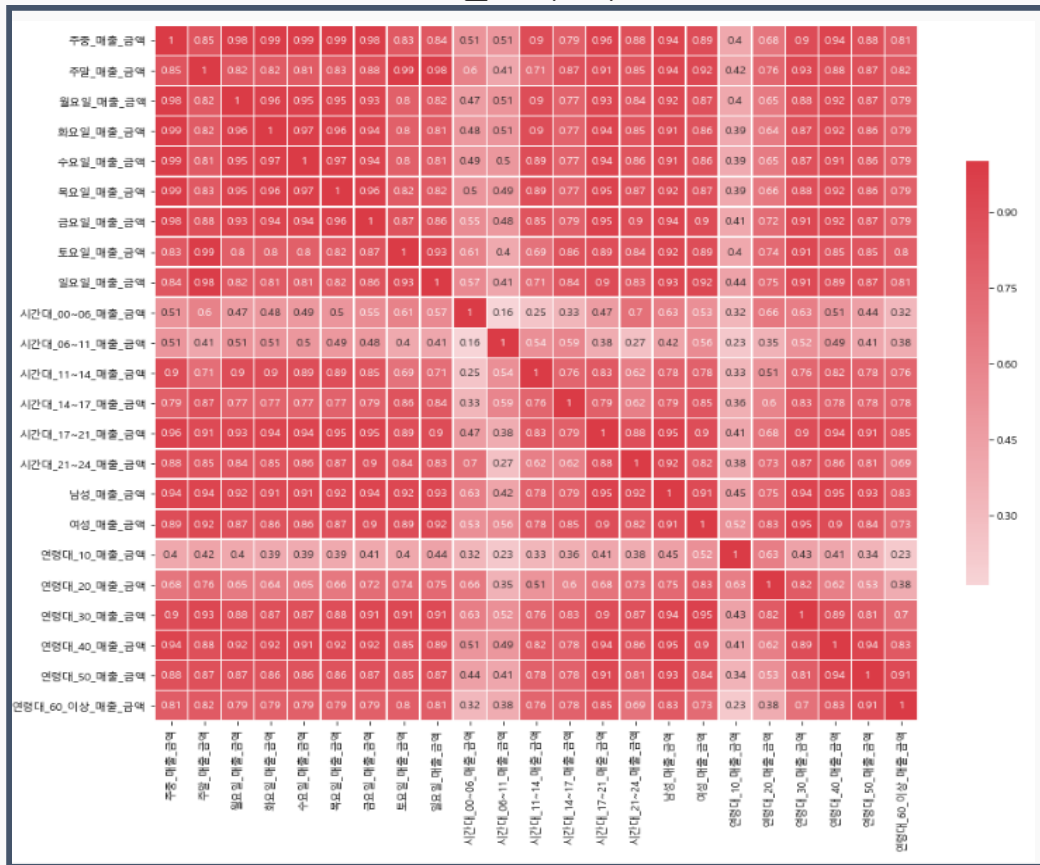
3) 주성분 별 가중치 산정 :
[0.22604656 0.17769522 0.09826107 0.06532068 0.05777666 0.05385568 0.05041156 0.04852136 0.0456105 0.04376407 0.04217785 0.03260387 0.0299764 0.02797851]

- ✓ 주성분 분석을 통해 생성된 지수의 성능은 차원 축소에 사용된 변수들 간 상관관계와 관련
 - ✓ 매출비율 지수 생성에 사용할 변수들의 상관계수 확인
- 서로 연관성 떨어져 지수간 상관관계 확인 후 사용 여부 결정

데이터 전처리

:차원축소-지수 생성과정

< 매출가격지수 >



```
scaler = StandardScaler()
scaled_price = scaler.fit_transform(market[price_columns])
```

```
pca2 = PCA(n_components=4)
pca2.fit(scaled_price)
```

```
# 누적 고유치 비율
pca2_price = list(pca2.explained_variance_ratio_)
print("1) pca성분 별 비율 : ")
print(pca2_price)
print("")
print("2) pca성분 별 비율의합 : ")
print(sum(pca2_price))
print("")
```

```
# 가중치 산정
pca2_weight = pca2_price / sum(pca2_price)
print("3) 가중치 산정 : ")
print(pca2_weight)
```

```
ypca2 = pca2.fit_transform(scaled_price)
```

```
market['price_index'] = (ypca2 * pca2_weight).sum(axis=1)
```

1) pca성분 별 비율 :
[0.7877639736315113, 0.06089212721428026, 0.042639473132818115, 0.030933063283907096]

2) pca성분 별 비율의합 :
0.9222286372625168

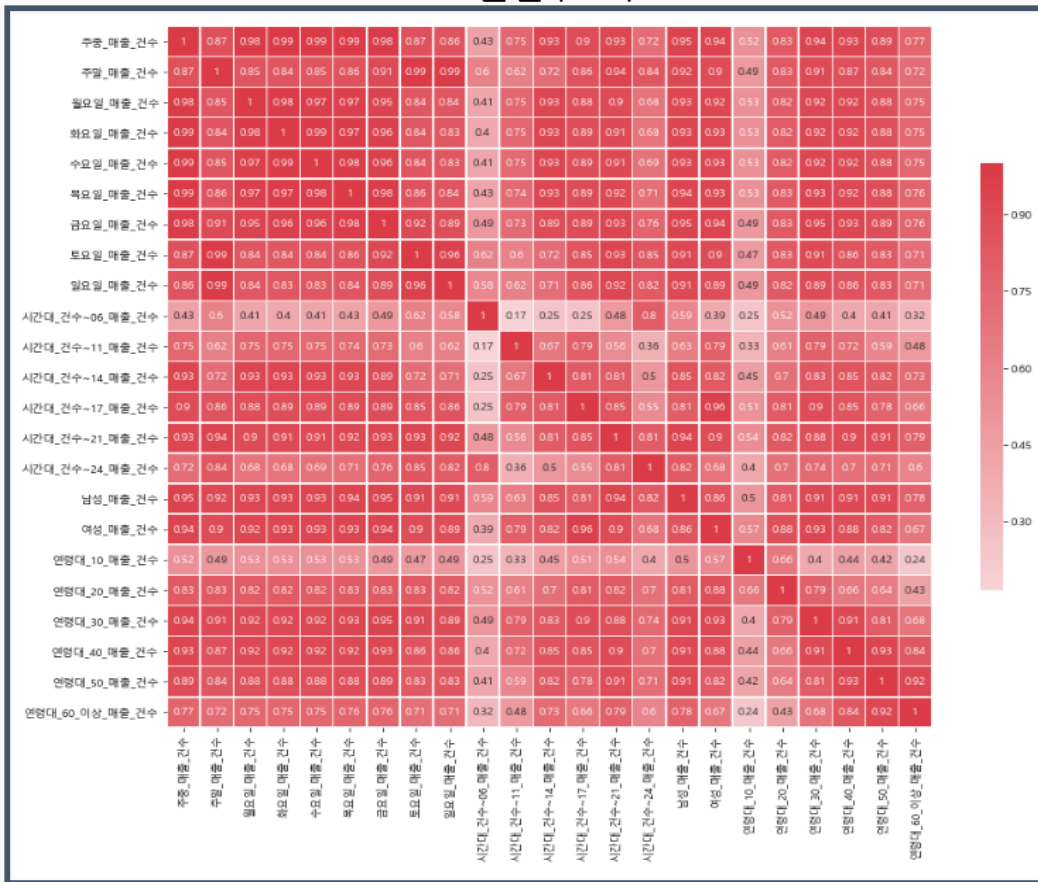
3) 가중치 산정 :
[0.85419596 0.06602715 0.04623525 0.03354164]

- ✓ 매출가격 지수생성에 사용할 변수들의 상관계수를 확인한 결과 서로 연관성이 존재
- ✓ 새벽시간 매출 금액과 10대 매출금액의 값은 다른 변수에 비해 값이 현저히 적어 상관관계가 낮게 도출

데이터 전처리

:차원축소-지수 생성과정

< 매출건수지수 >



```
scaler = StandardScaler()
scaled_count = scaler.fit_transform(market[count_columns])
```

```
pca3 = PCA(n_components=4)
pca3.fit(scaled_count)
```

```
# 누적 고유치 비율
pca3_count = list(pca3.explained_variance_ratio_)
print("1) pca성분 별 비율")
print(pca3_count)
print("")
print("2) pca성분 별 비율의합")
print(sum(pca3_count))
print("")
# 가중치 산정
pca3_weight = pca3_count / sum(pca3_count)
print("3) 가중치 산정")
print(pca3_weight)
```

```
ypca3 = pca3.fit_transform(scaled_count)
```

```
market['count_index'] = (ypca3 * pca3_weight).sum(axis=1)
```

1) pca성분 별 비율
[0.8071307705811224, 0.06352930246578004, 0.04389689594740322, 0.027705473760926122]

2) pca성분 별 비율의합
0.9422624427552317

3) 가중치 산정
[0.85658807 0.06742209 0.0465867 0.02940314]



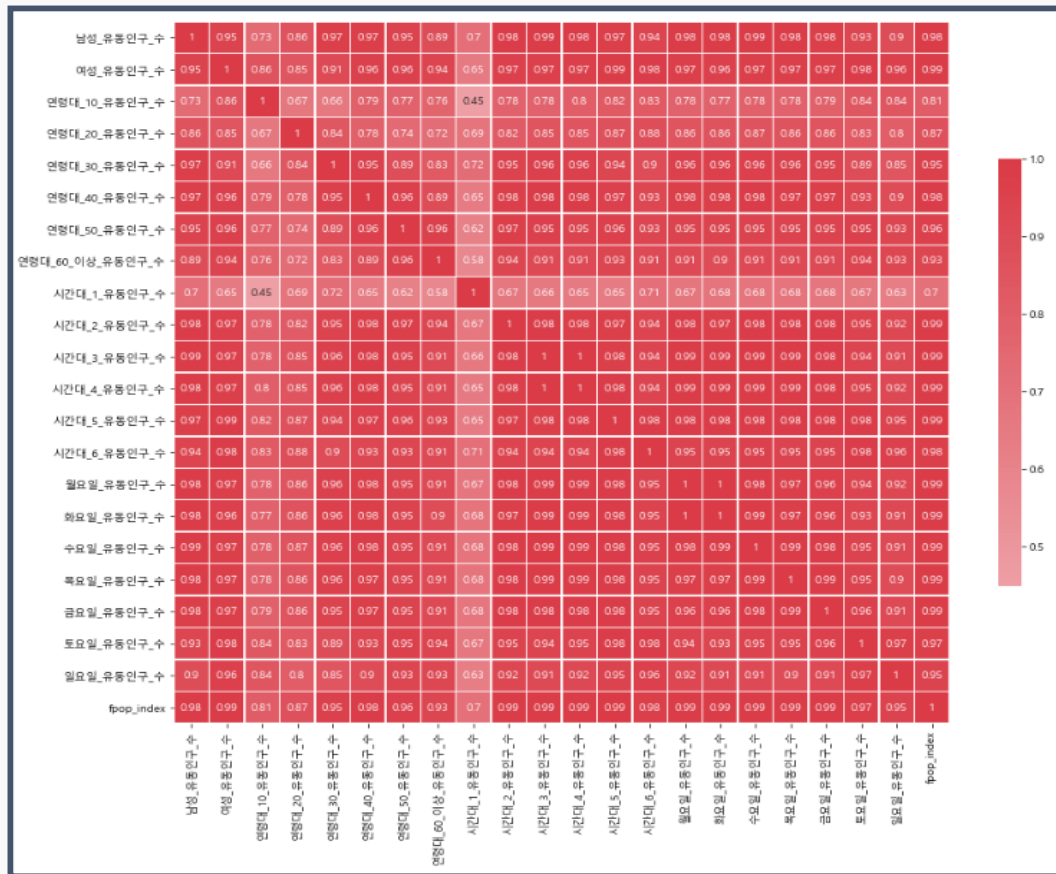
매출건수 지수생성에 사용할 변수들의 상관계수를 확인한 결과 서로 연관성이 존재

새벽시간 매출 건수와 10대 매출건수의 값은 다른 변수에 비해 값이 현저히 적어 상관관계가 낮게 도출

데이터 전처리

:차원축소-지수 생성과정

<유동인구지수>



```
# StandardScaler (평균을 0 으로 만들어 줌)
scaler = StandardScaler()
scaled_fpop = scaler.fit_transform(market[fpop_columns])
```

```
pca4 =PCA(n_components=2)
pca4.fit(scaled_fpop)
```

```
# 누적 고유치 비율
pca4_fpop = list(pca4.explained_variance_ratio_)
print("1) pca성분 별 비율")
print(pca4_fpop)
print("")
print("2) pca성분 별 비율의합")
print(sum(pca4_fpop))
print("")
```

```
# 가중치 산정
pca4_weight = pca4_fpop / sum(pca4_fpop)
print("3) 가중치 산출")
print(pca4_weight)
```

```
ypca4 = pca4.fit_transform(scaled_fpop)
```

```
market['fpop_index'] = (ypca4 * pca4_weight).sum(axis=1)
```

1) pca성분 별 비율
[0.9134766657353719, 0.031268867921342015]

2) pca성분 별 비율의합
0.9447455336567139

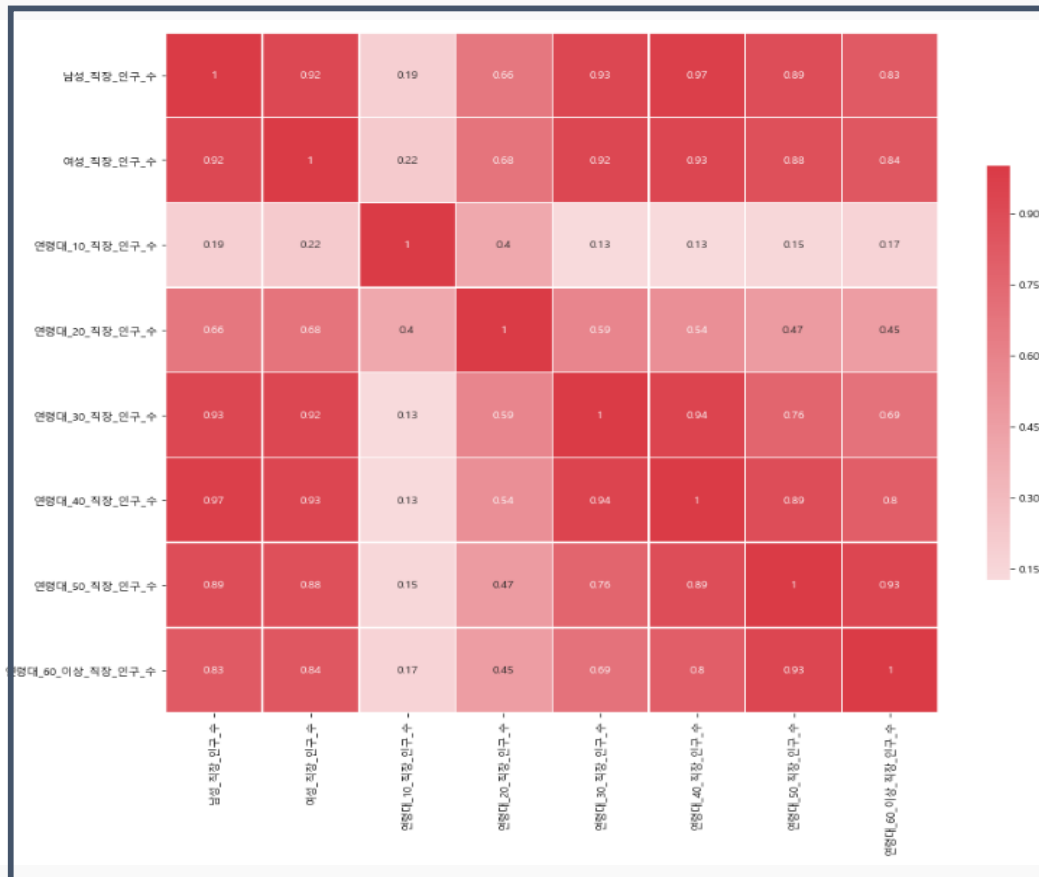
3) 가중치 산출
[0.96690234 0.03309766]

- ✓ 유동인구 지수생성에 사용할 변수들의 상관계수를 확인한 결과 서로 연관성이 존재
- ✓ 새벽 시간 유동인구수는 다른 변수에 비해 값이 현저히 적어 상관관계가 낮게 도출

데이터 전처리

:차원축소-지수 생성과정

<직장인구지수>



```
# StandardScaler (평균을 0 으로 만들어 줌)
scaler = StandardScaler()
scaled_work_pop = scaler.fit_transform(market[work_pop_columns])

pca5 =PCA(n_components=3)
pca5.fit(scaled_work_pop)

# 누적 고유치 비율
pca5_work_pop = list(pca5.explained_variance_ratio_)
print("1) pca성분 별 비율")
print(pca5_work_pop)
print("")
print("2) pca성분 별 비율의합")
print(sum(pca5_work_pop))
print("")
# 가중치 산정
pca5_weight = pca5_work_pop / sum(pca5_work_pop)
print("3) 가중치 산정")
print(pca5_weight)

1) pca성분 별 비율
[0.7589033786934046, 0.1246488075068246, 0.06187306575811388]

2) pca성분 별 비율의합
0.9454252519583431

3) 가중치 산정
[0.80271114 0.13184417 0.06544469]
```

- ✓ 직장인구 지수생성에 사용할 변수들의 상관계수를 확인한 결과 서로 연관성이 존재
- ✓ 10대 직장인구수는 다른 변수에 비해 값이 현저히 적어 상관관계가 낮게 도출

데이터 전처리

:차원축소-생성된지수의검정



생성된 지수와 지수를 만드는데 사용된 변수들과의 상관관계를 확인하여 지수의 성능을 검증



매출비율 지수는 변수들의 특성을 충분히 반영하지 못하기 때문에 분석에서 제외

데이터 전처리

: 차원 축소 - 생성된 지수의 검정

price_index	count_index	fpop_index	work_pop_index
0.989109	1.140346	6.024815	-1.601738
-1.530652	-1.690629	6.024815	-1.601738
-1.816216	-1.876242	6.024815	-1.601738
-1.384527	-0.800722	6.024815	-1.601738
1.790485	5.934934	6.024815	-1.601738

MinMaxScaler

price_index	count_index	fpop_index	work_pop_index
0.015188	0.009286	0.191638	0.026806
0.003322	0.001791	0.191638	0.026806
0.001977	0.001300	0.191638	0.026806
0.004010	0.004147	0.191638	0.026806
0.018962	0.021980	0.191638	0.026806



주성분 분석을 통해 생성된 지수 값 중 음수가 존재



Sklearn패키지의 MinMaxScaler 를 사용하여,
0과 1사이의 값으로 스케일을 조정

데이터 전처리

:파생변수 생성 및 변환

변수 생성 방법	설명
<pre>market_df['점포별평균당월매출금액'] = market_df['당월_매출_금액']/market_df['점포'] market_df['점포별평균당월매출건수'] = market_df['당월_매출_건수']/market_df['점포']</pre>	점포 1개 당 평균 금액, 건수
<pre>market_df['아파트_가구_비율'] = market_df['아파트_가구_수'] / market_df['총_가구_수'] market_df['비_아파트_가구_비율'] = market_df['비_아파트_가구_수'] / market_df['총_가구_수']</pre>	총 가구수에 대한 아파트 가구와 비 아파트 가구 비율 확인
<pre>market['년도'] = market['기준_년월_코드'].apply(lambda x:str(x)[:4]).astype(int) market['월'] = market['기준_년월_코드'].apply(lambda x:str(x)[4:]).astype(int)</pre>	연별 파악, 월별 파악을 위해 기준 년월을 년과 월로 분리

데이터 전처리

:파생변수 생성 및 변환

변수 생성 방법	설명
<pre># 당월 매출 금액 변화량 data = sales.pivot_table("당월_매출_금액", "기준_년월_코드", "상권_코드_명") data = data.diff() data = pd.DataFrame(data.unstack()).pivot_table(0, index=['기준_년월_코드', '상권_코드_명']) data = data.reset_index() market_df = pd.merge(market_df, data, on=['기준_년월_코드', '상권_코드_명'], how='left') market_df.rename(columns={0: "전월_대비_매출_금액_변화량"}, inplace=True)</pre>	전월 대비 해당 월의 매출 금액 변화량
<pre># 당월 매출 건수 변화량 data1 = sales.pivot_table("당월_매출_건수", "기준_년월_코드", "상권_코드_명") data1 = data1.diff() data1 = pd.DataFrame(data1.unstack()).pivot_table(0, index=['기준_년월_코드', '상권_코드_명']) data1 = data1.reset_index() market_df = pd.merge(market_df, data1, on=['기준_년월_코드', '상권_코드_명'], how='left') market_df.rename(columns={0: "전월_대비_매출_건수_변화량"}, inplace=True)</pre>	매출 건수 변화량
<pre># 유동인구수 변화량 data2 = com_fpop.pivot_table("총_유동인구_수", "기준_년월_코드", "상권_코드_명") data2 = data2.diff() data2 = pd.DataFrame(data2.unstack()).pivot_table(0, index=['기준_년월_코드', '상권_코드_명']) data2 = data2.reset_index() market_df = pd.merge(market_df, data2, on=['기준_년월_코드', '상권_코드_명'], how='left') market_df.rename(columns={0: "전월_대비_유동인구수_변화량"}, inplace=True)</pre>	총 유동인구 변화량

데이터 전처리

:상권별 군집화

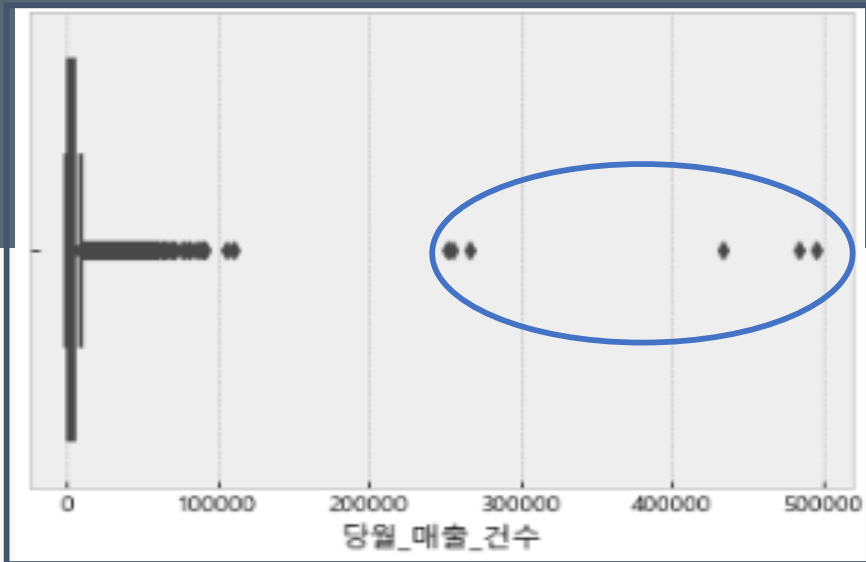
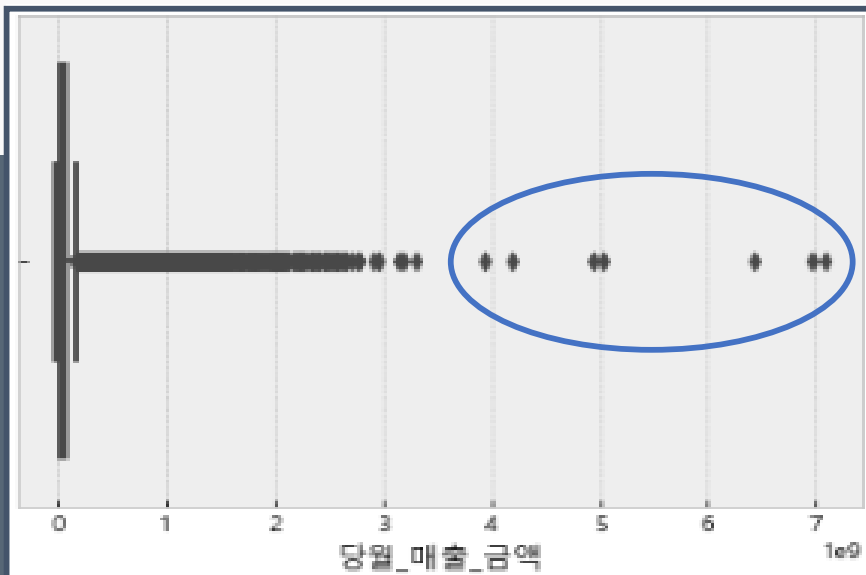
	기준_년월_코드	상권_코드	상권_코드_명	서비스_업종_코드	서비스_업종_코드_명	당월_매출_금액	당월_매출_건수	점포	총_상주_인구_수	총_가구_수	아파트_가구_수	비_아파트_가구_수
0	201501	11947	당산로44길	CS100001	한식음식점	96722641	4527	2	194.0	84.0	84.0	0.0
1	201501	11947	당산로44길	CS100004	양식집	14242932	910	1	194.0	84.0	84.0	0.0
2	201501	11947	당산로44길	CS100005	분식집	5084735	653	1	194.0	84.0	84.0	0.0
3	201501	11947	당산로44길	CS100006	패스트푸드점	18820989	2098	2	194.0	84.0	84.0	0.0
4	201501	11947	당산로44길	CS100008	제과점	124446237	11102	2	194.0	84.0	84.0	0.0

✓ 각 상권의 업종별 특성 고려하여 비슷한 유형끼리 분류. 효율적 분석을 위해 K-means 군집분석 진행

✓ 군집별로 어떤 특성을 가지고 있는지 파악

데이터 전처리

:상권별 군집화- 이상치처리

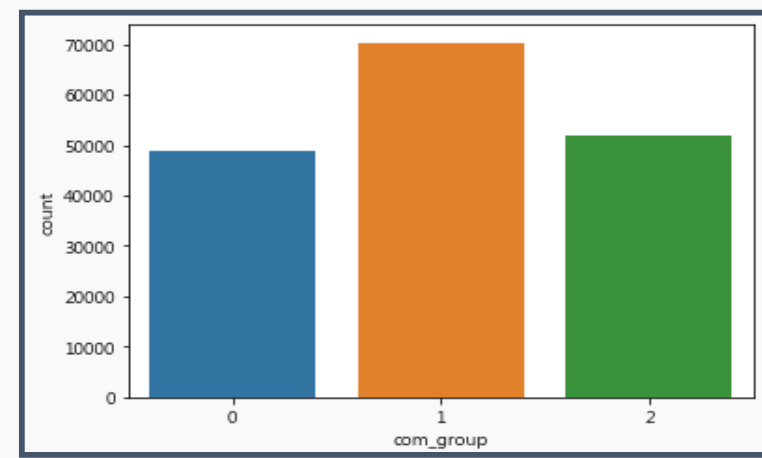
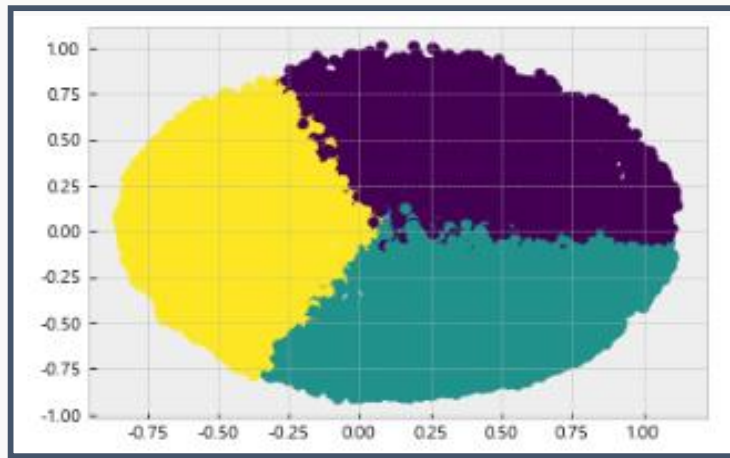
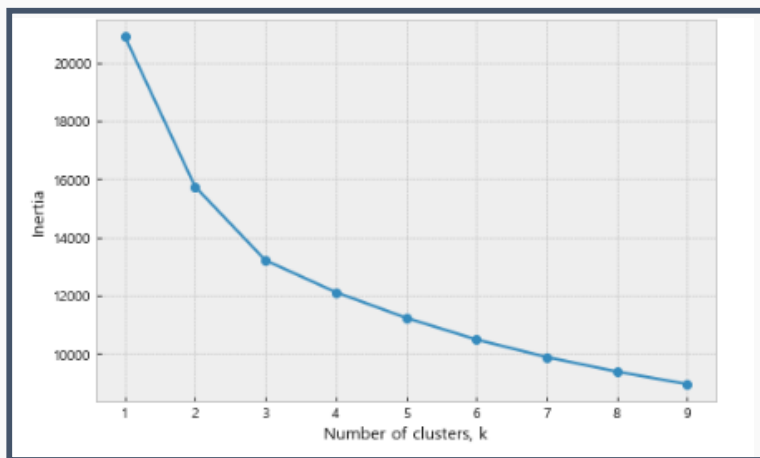


- ✓ K-means 군집분석은 이상치에 민감하기 때문에 이상치를 제거하고 분석
- ✓ 이상치 제거 기준

- 1) 다른 값들에 비해 지나치게 높은 값이거나 갑자기 급격히 상승한 경우
- 2) 오류 값을 확인한 경우
- 3) 정규화 했음에도 너무 값의 분포가 동떨어져 있는 경우

데이터 전처리

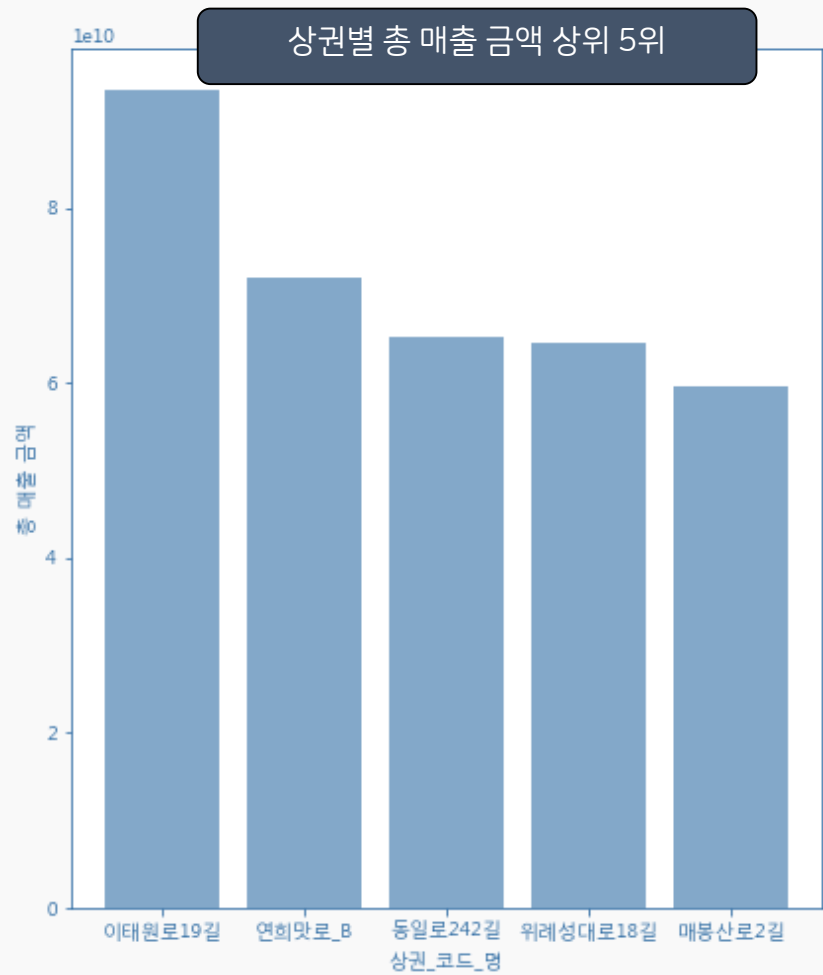
:상관별 군집화- 이상치처리



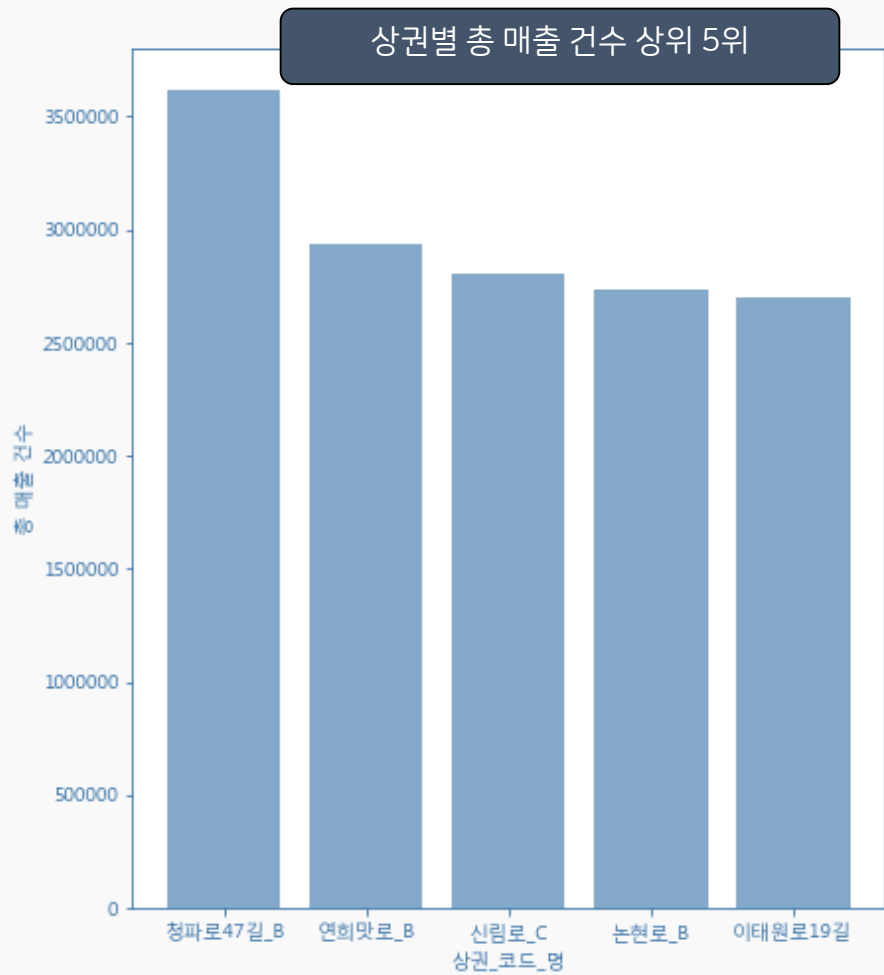
- ✓ 거리 계산을 위해 sklearn패키지의 Normailze, StandardScaler 모듈을 통해 정규화와 z-score 표준화
- ✓ 고차원의 데이터를 주성분 분석을 통해 2차원 공간에 투영
- ✓ Scree Plot을 통해 알맞은 군집의 갯수 선택하여 3개의 군집으로 유형 분류

데이터 분석 및 시각화

2년간 상권별 매출 금액/건수 상위 5위



1위 : 이태원로19길

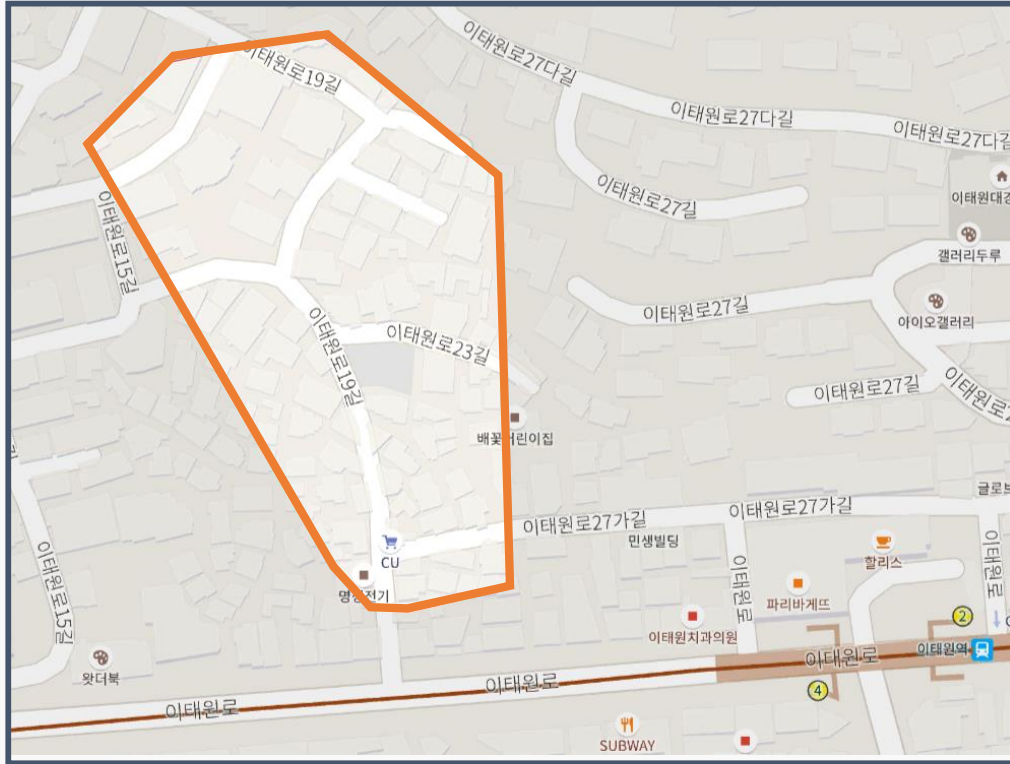


1위 : 청파로47길_B

데이터 분석 및 시각화

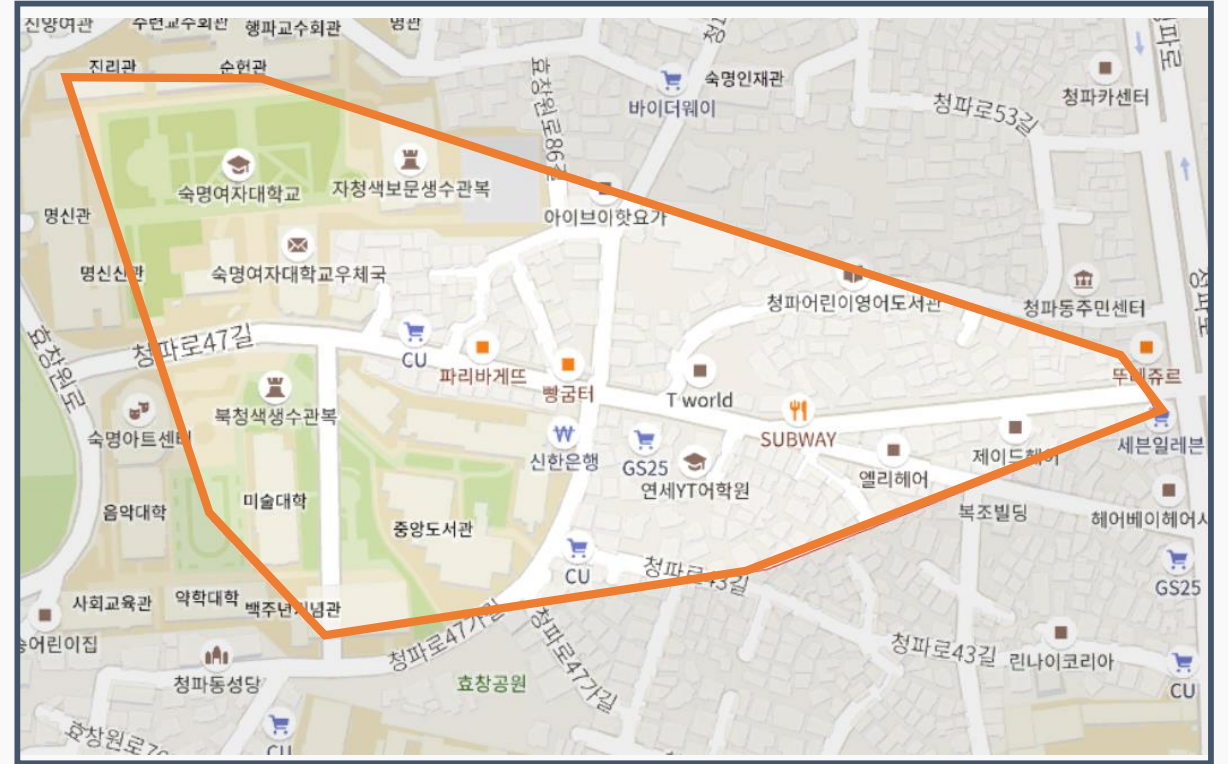
상위 상권 위치

이태원로19길



✓ 이태원역과 가깝고 음식점이 밀집되어 있는 지역

청파로47길



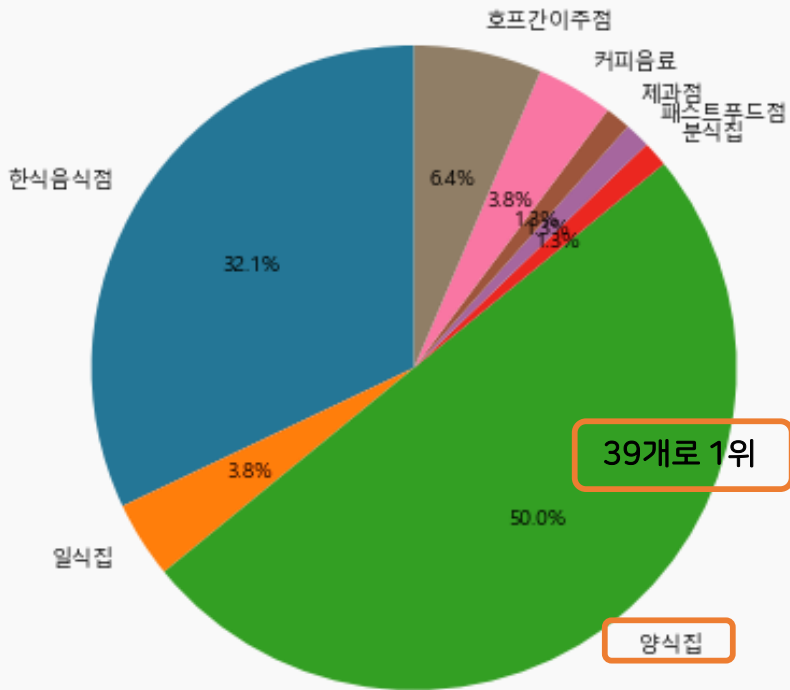
✓ 숙명여대 앞이므로 대학가 특수성이 있음

데이터 분석 및 시각화

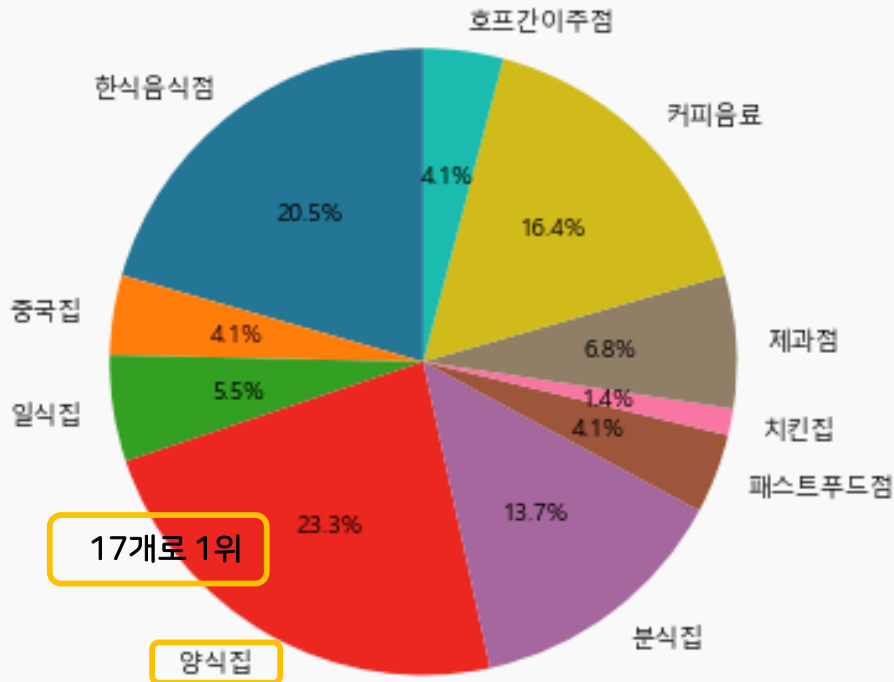
상위 상권 업종별 점포 비율

2016년 12월 기준

이태원로19길 업종별 점포 비율



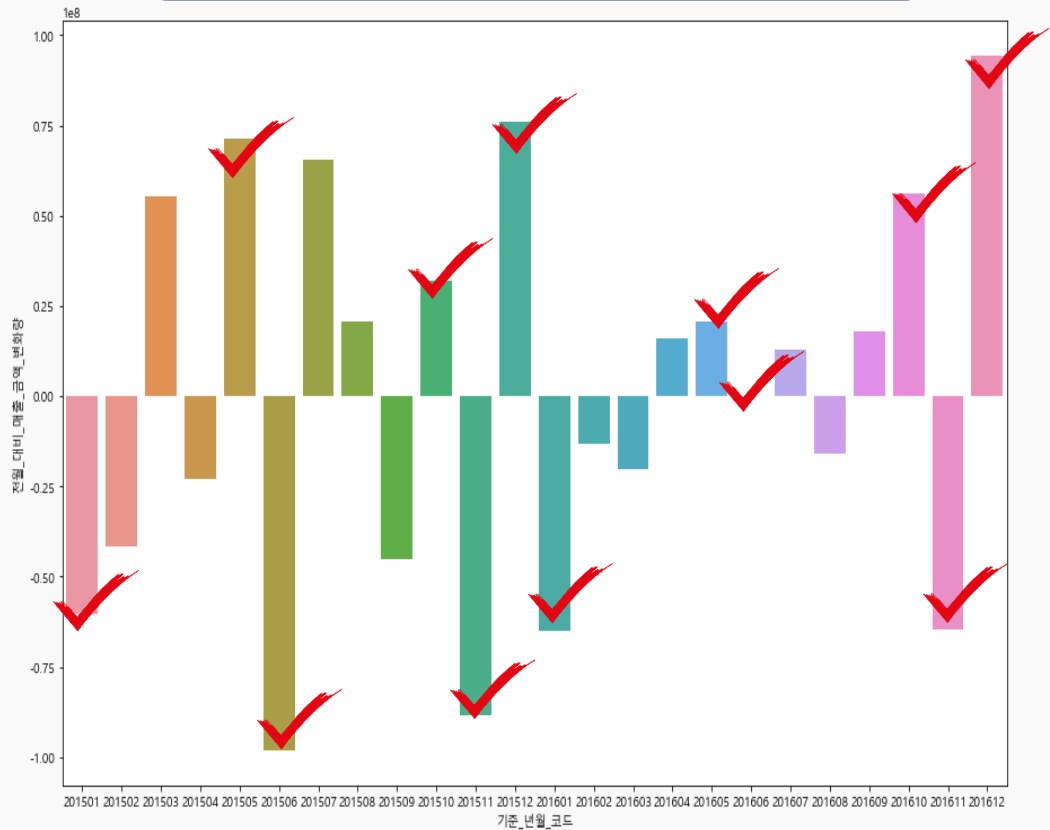
청파로47길_B 업종별 점포 비율



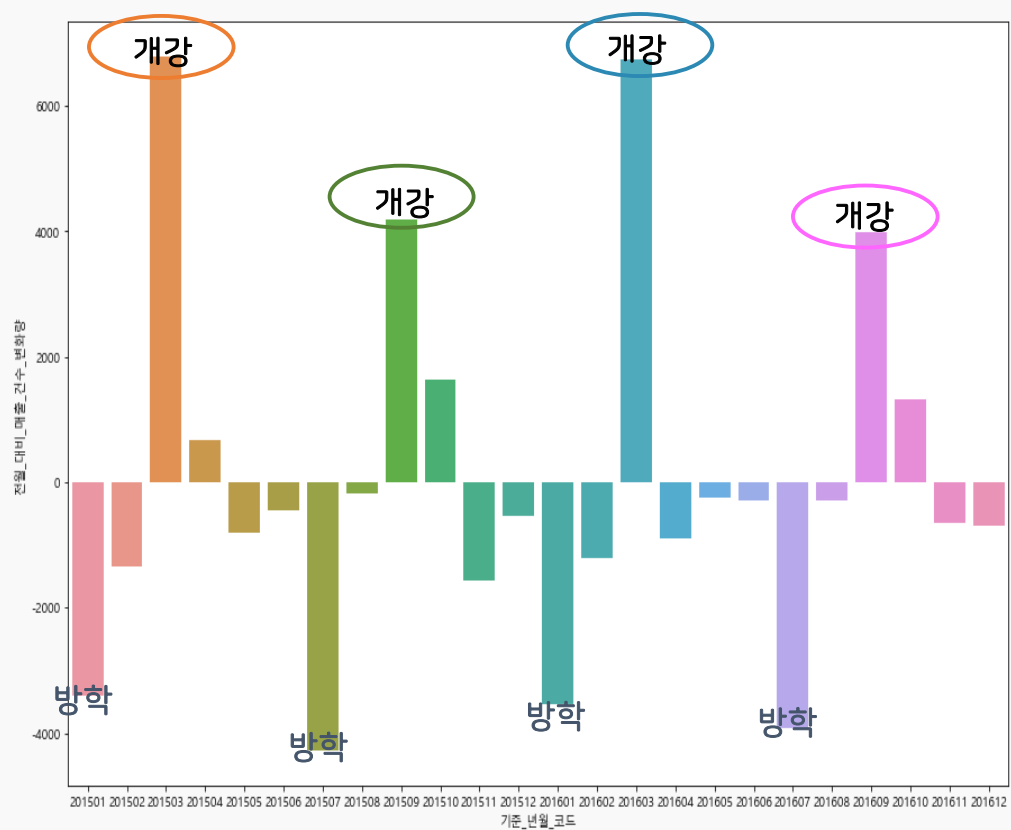
데이터 분석 및 시각화

전월 대비 매출 금액/ 건수 변화 파악

이태원로19길 전월 대비 매출 금액 변화량

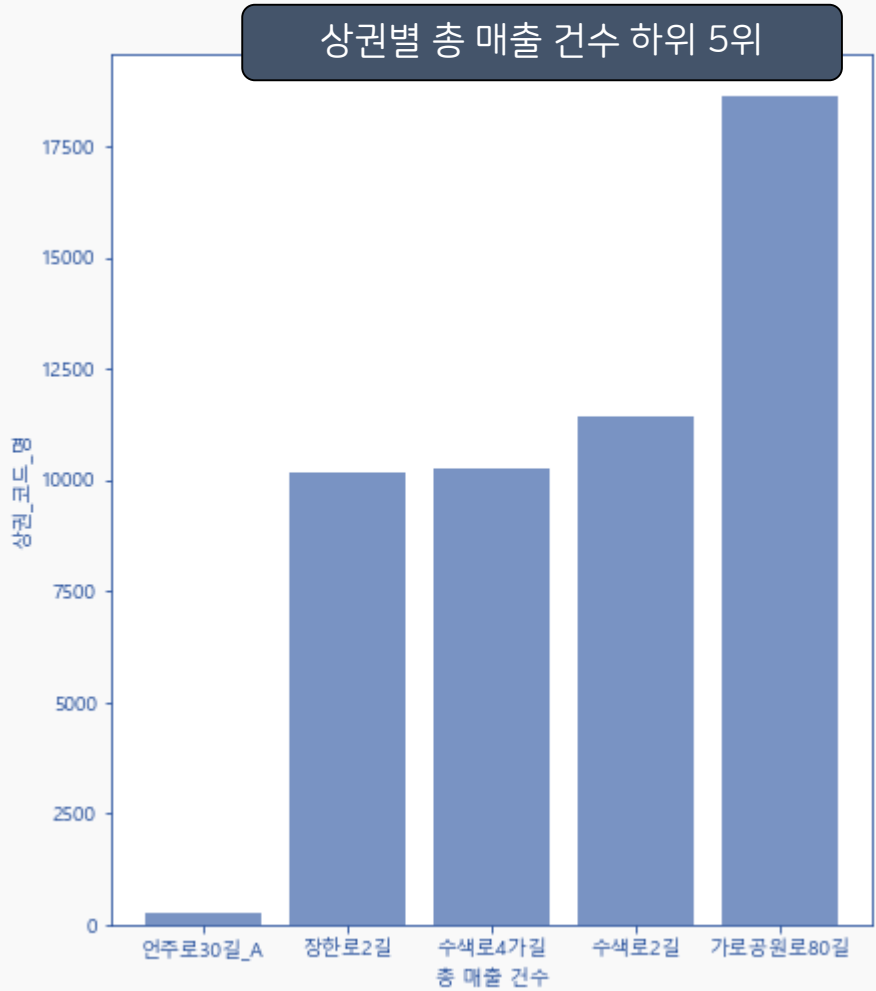
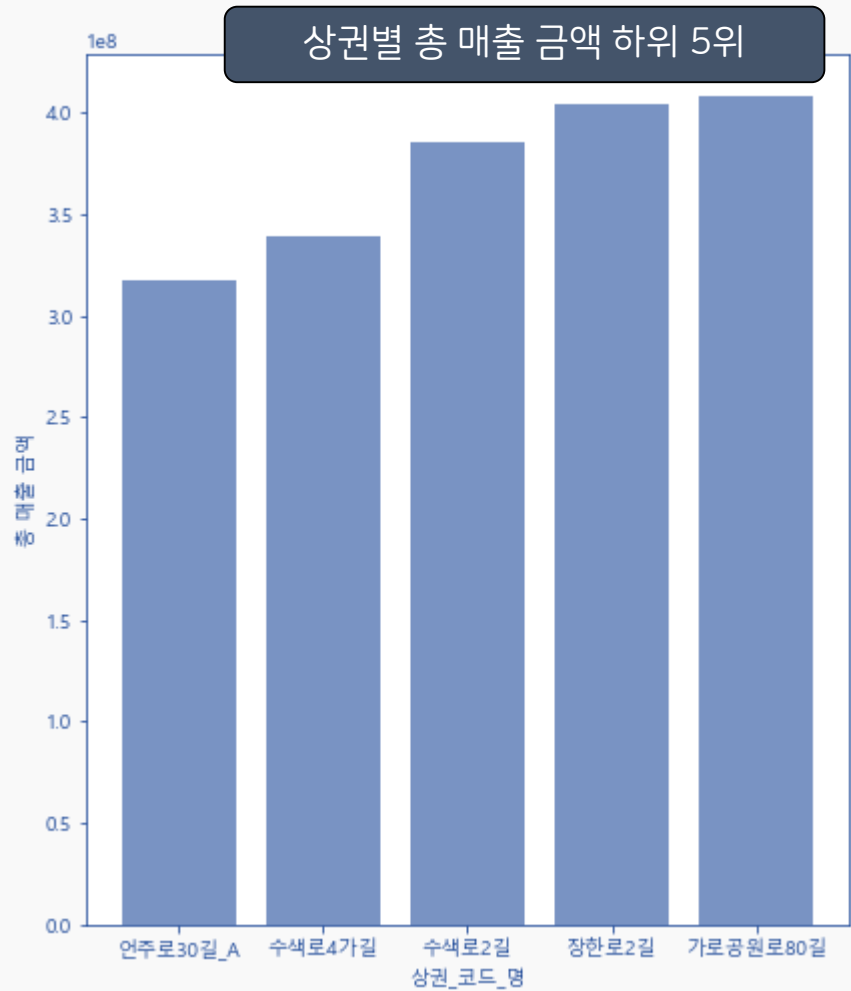


청파로47길_B 전월 대비 매출 건수 변화량



데이터 분석 및 시각화

2년간 상권별 매출 금액/건수 하위 5위



1위 : 연주로30길_A

데이터 분석 및 시각화

하위 상권 위치

언주로30길

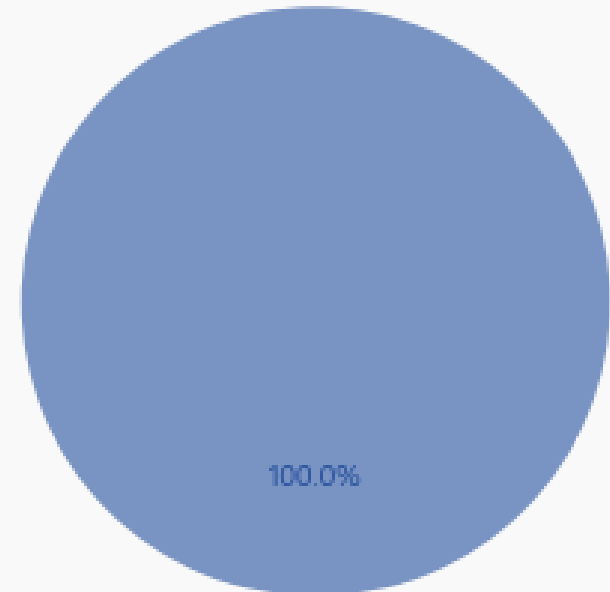


도곡역과 가깝고 고급 아파트가 밀집되어 있는 지역

하위 상권 업종 비율

2015년 10월 기준

언주로30길_A 상권의 업종별 점포 비율



호프간이주점

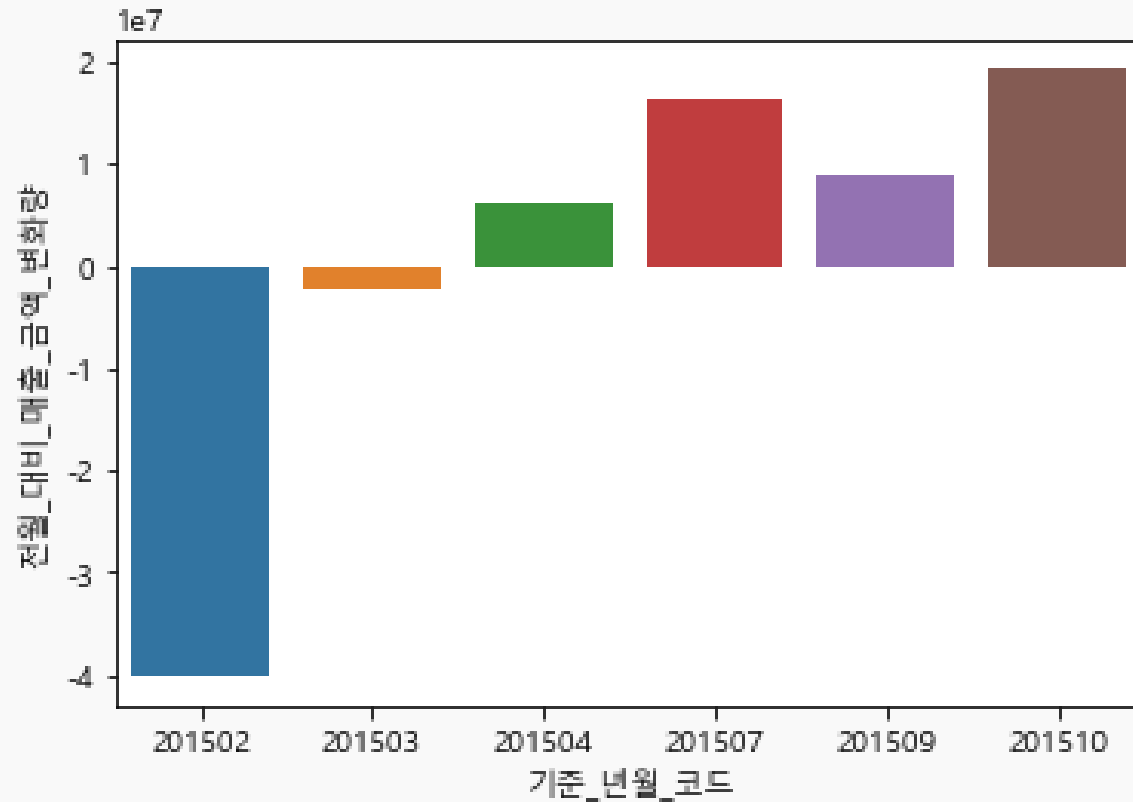


호프간이주점 1개

데이터 분석 및 시각화

전월 대비 매출 금액/건수 변화 파악

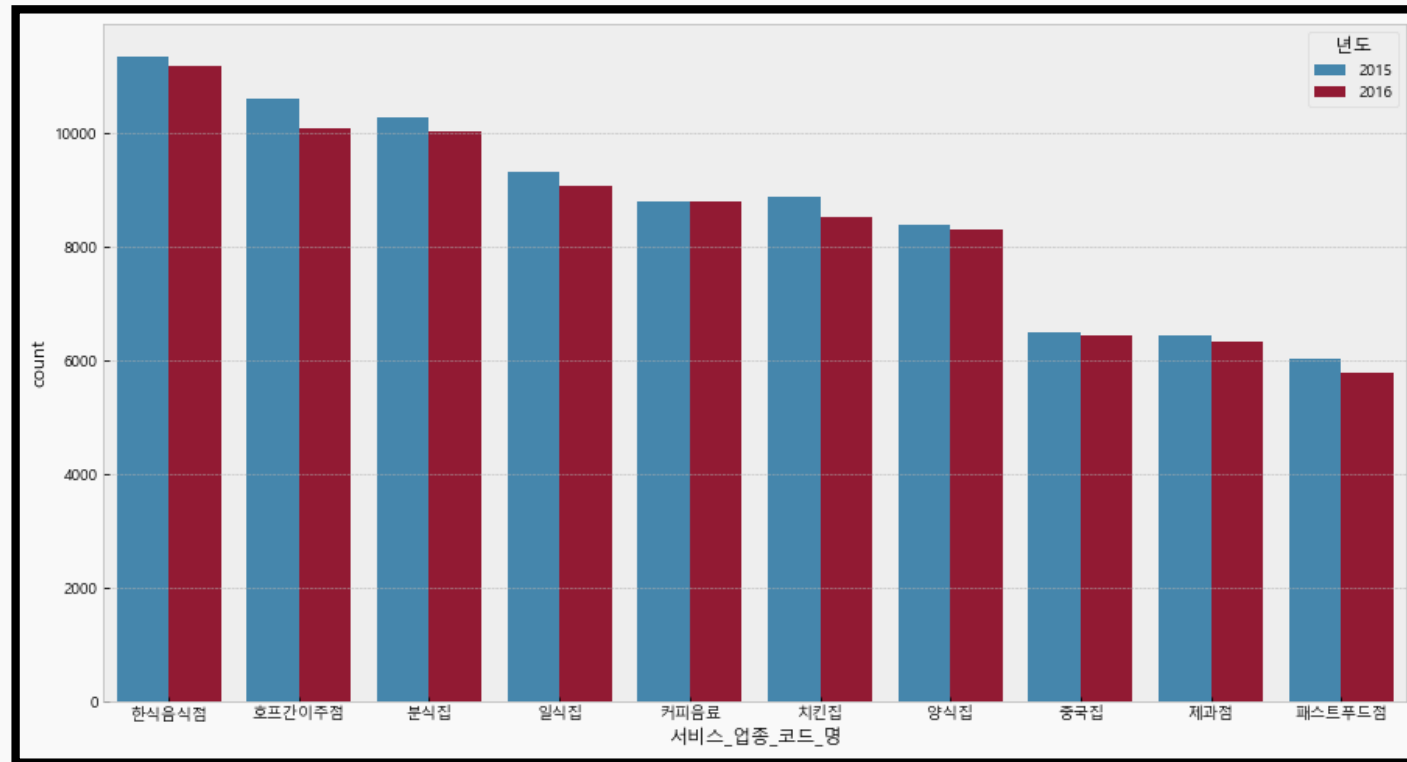
언주로30길_A 전월 대비 매출 금액 변화량



2015년 11월부터의 매출 금액 데이터가 존재하지 않음

데이터 분석 및 시각화

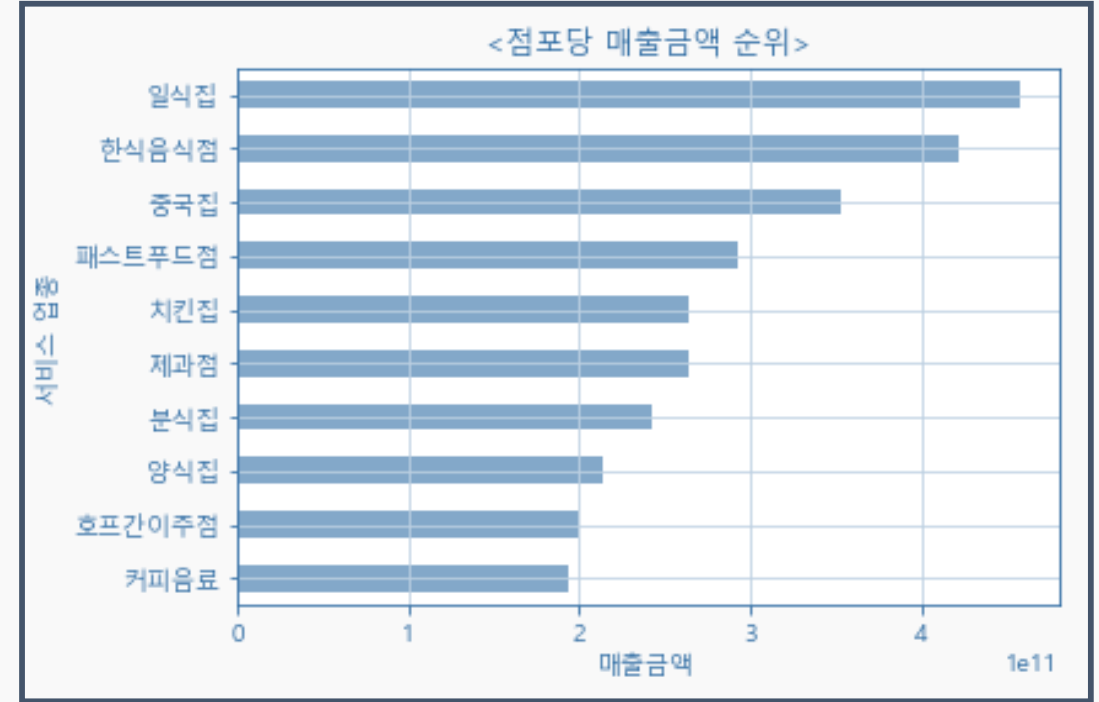
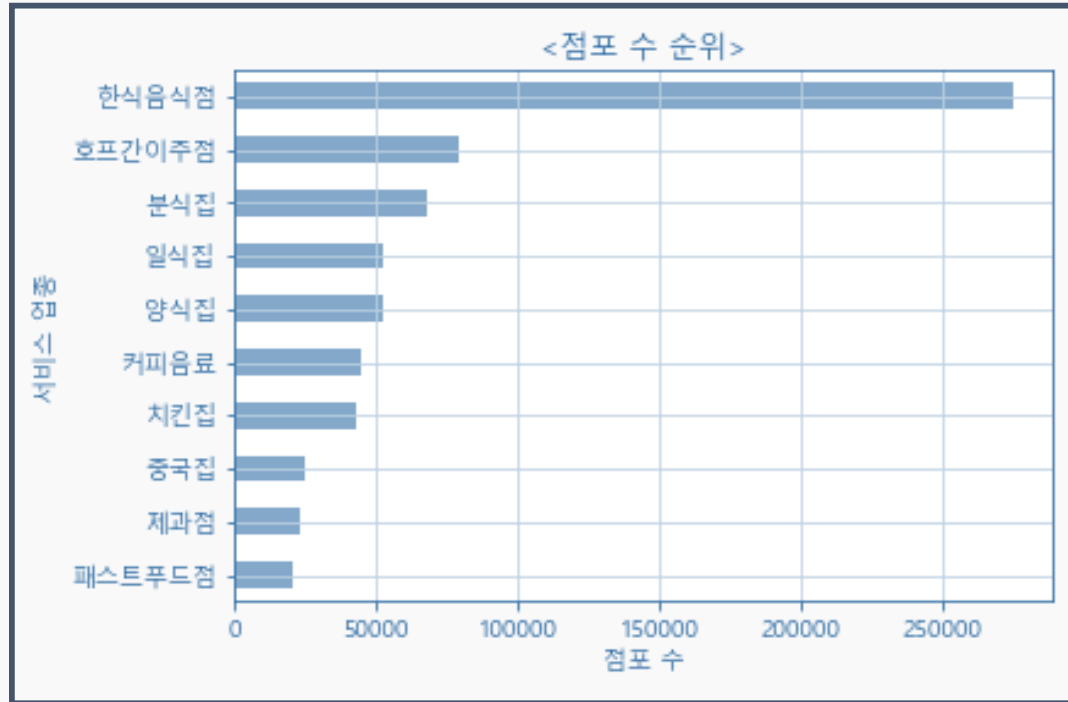
2년간 요식업 점포 수 변화



전반적으로 점포 수가 비슷하거나 감소

데이터 분석 및 시각화

점포 수 vs 매출금액지수

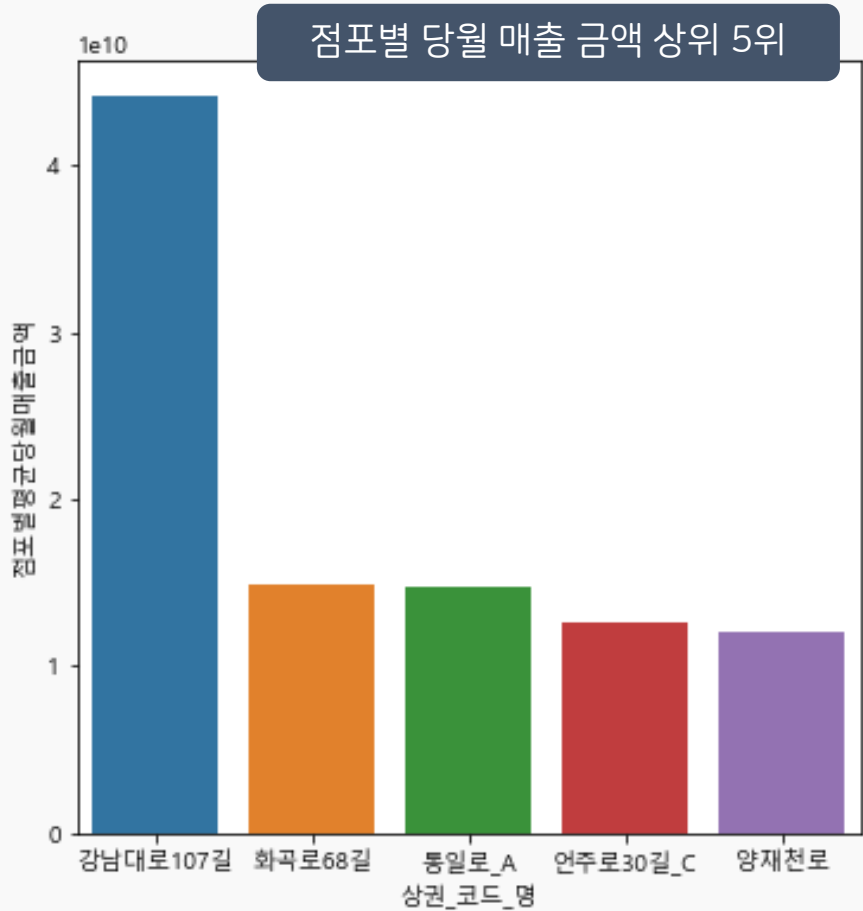


✓ 점포 수가 가장 많은 업종은 한식음식점

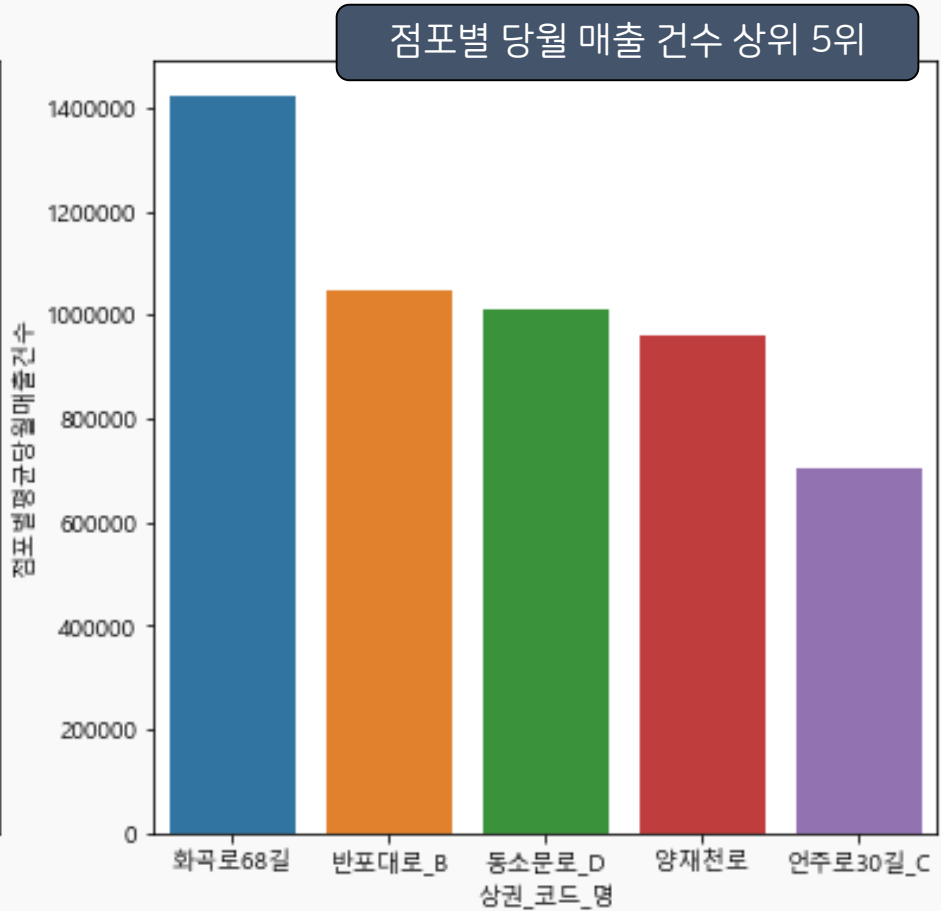
✓ 매출금액이 가장 높은 업종은 일식집

데이터 분석 및 시각화

점포별 당월 매출 금액/건수 상위 5위



1위 : 강남대로107길

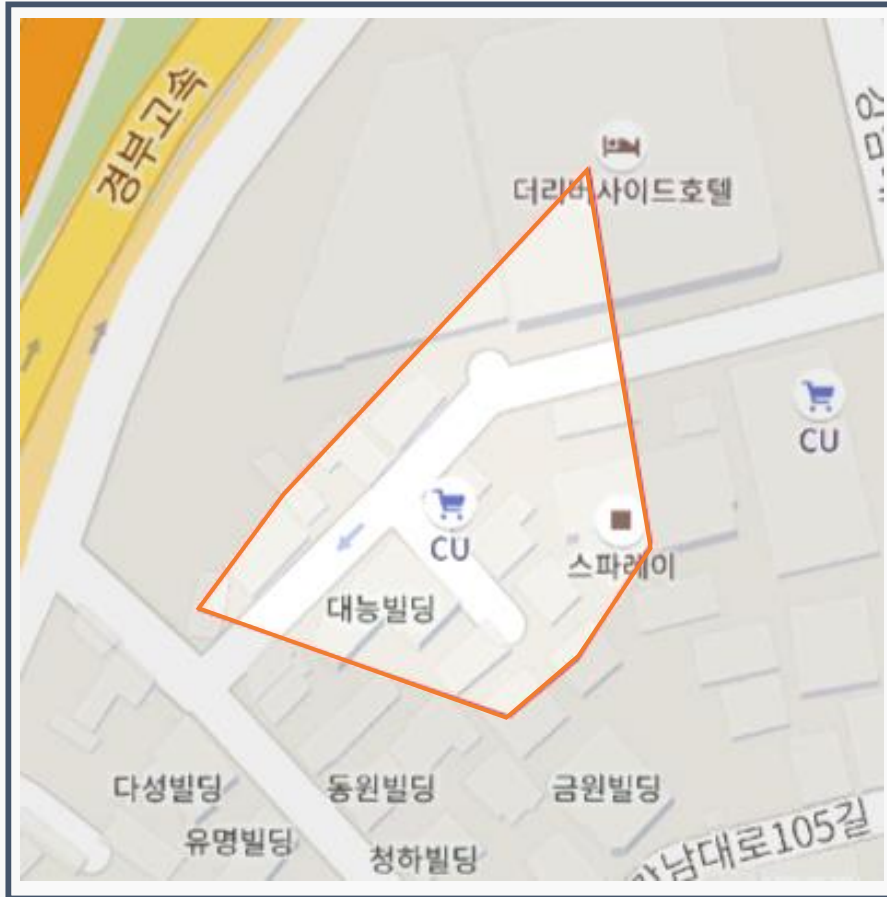


1위 : 화곡로68길

데이터 분석 및 시각화

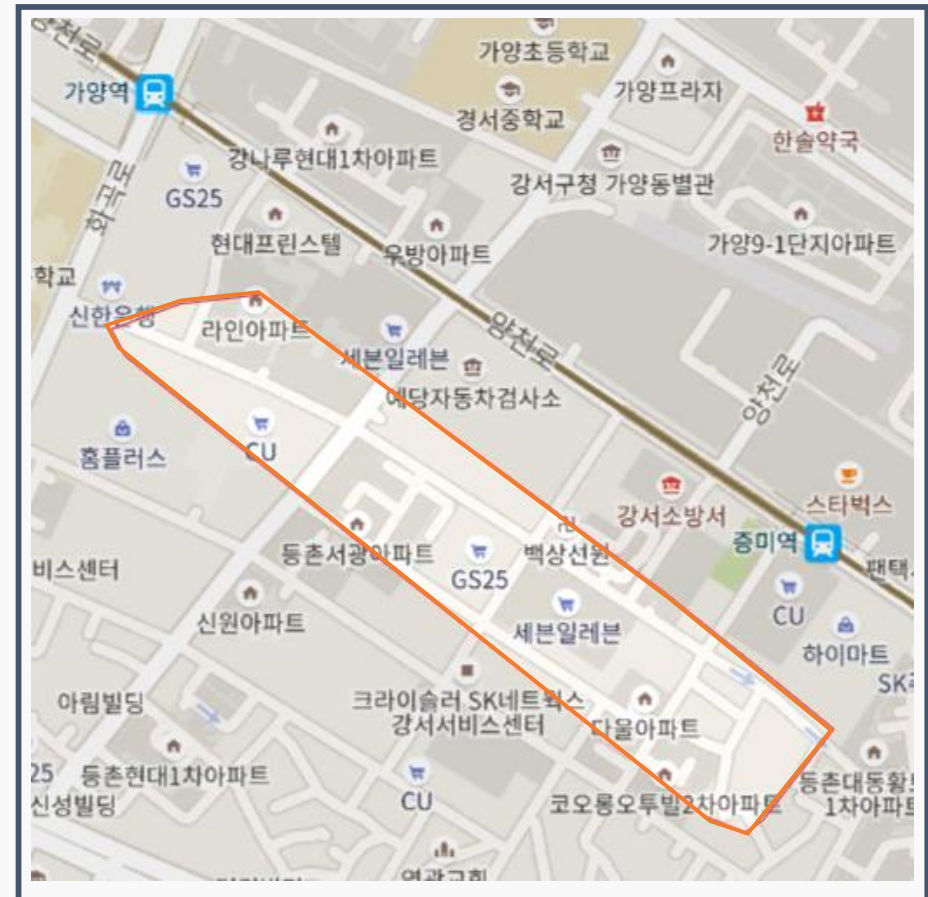
상위 상권 위치

강남대로107길



더리버사이드호텔이 위치해 있음

화곡로68길



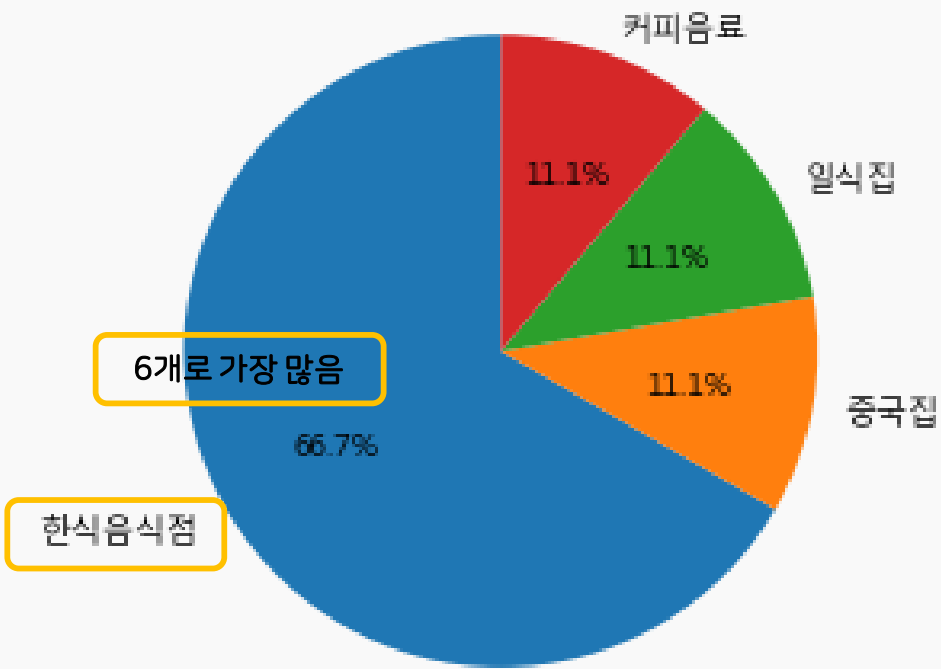
아파트 단지

데이터 분석 및 시각화

상위 상권 업종별 점포 비율

2016년 11월 기준

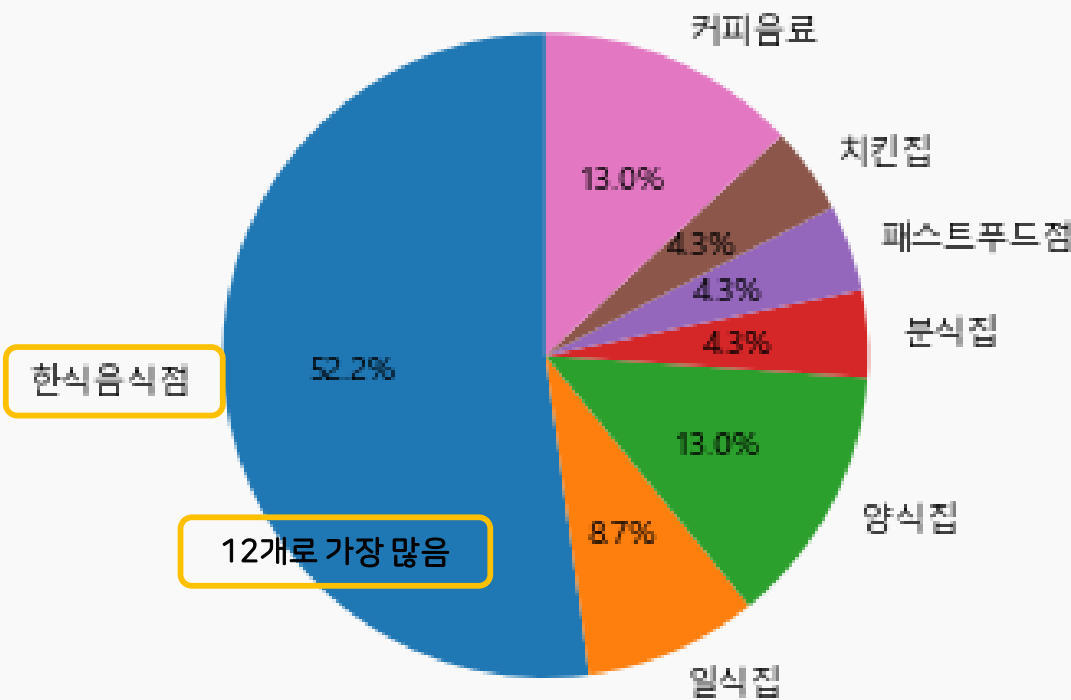
강남대로107길 업종별 점포 비율



한식음식점 6개

2016년 12월 기준

화곡로68길 업종별 점포 비율



데이터 분석 및 시각화

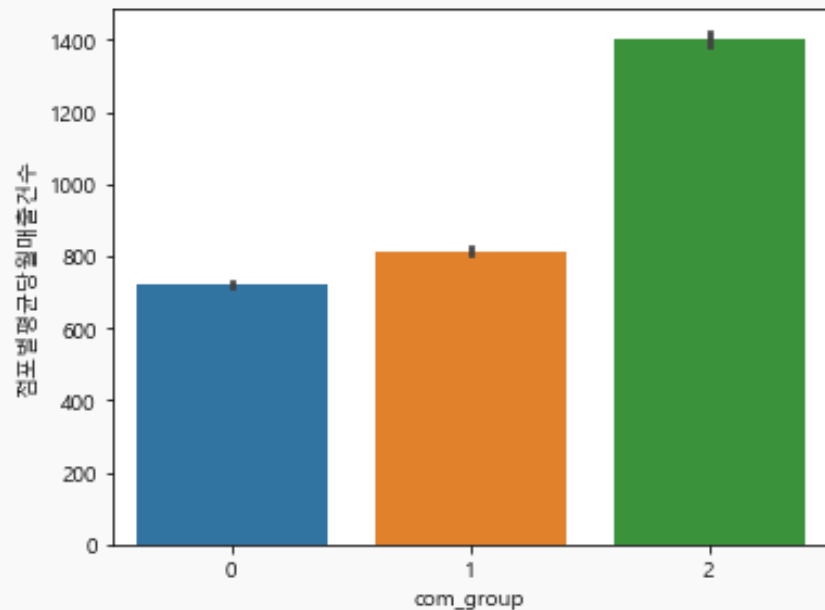
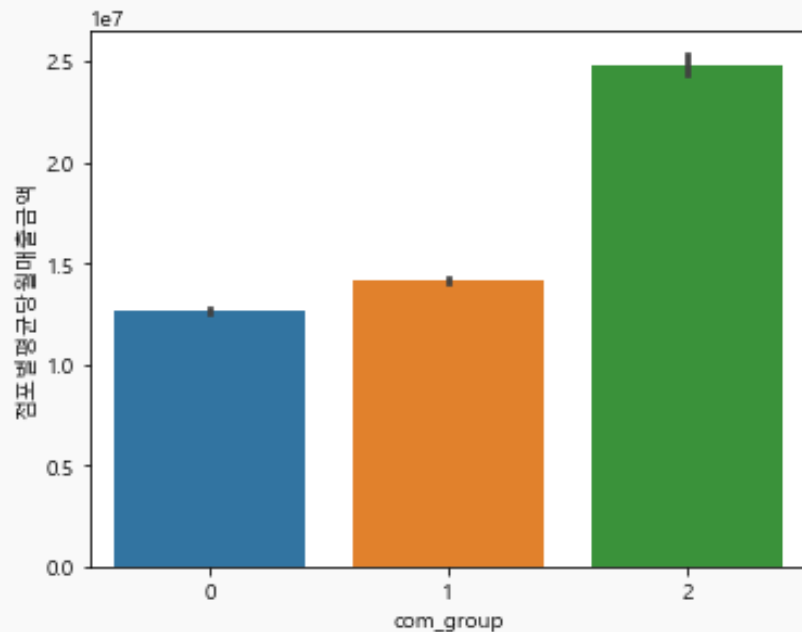
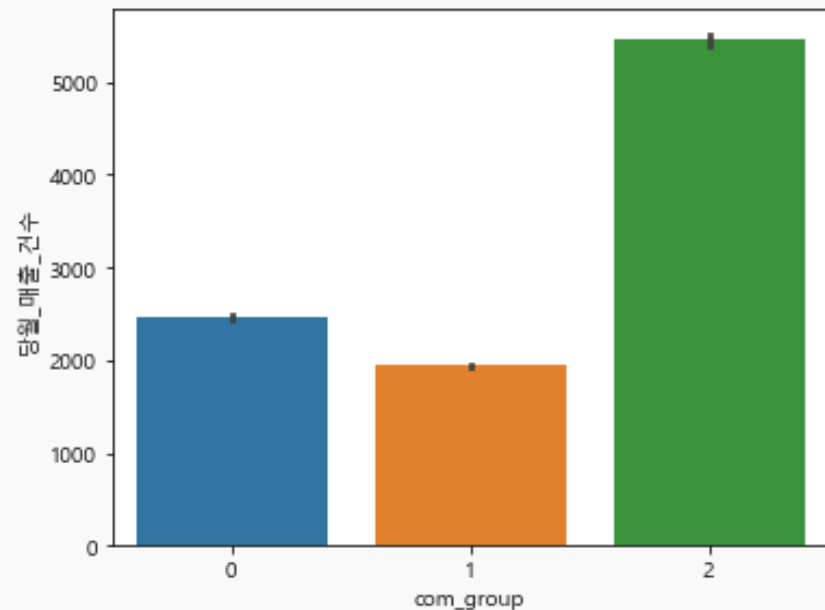
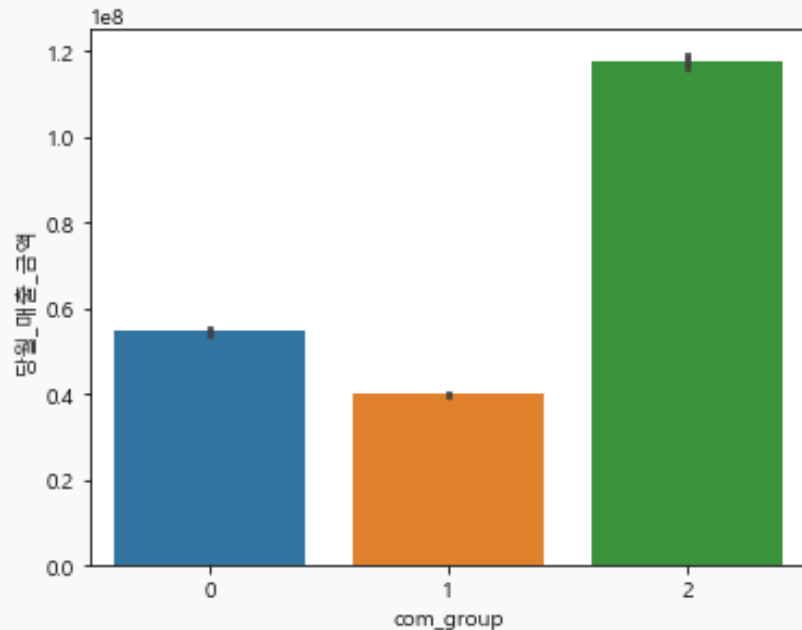
: 군집별 특성 파악



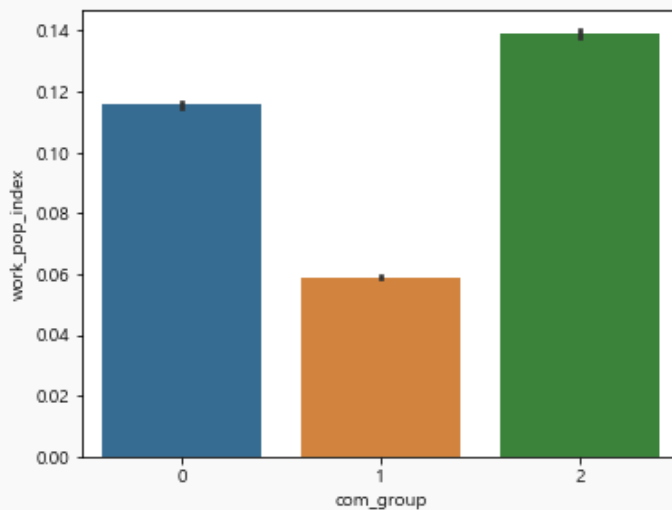
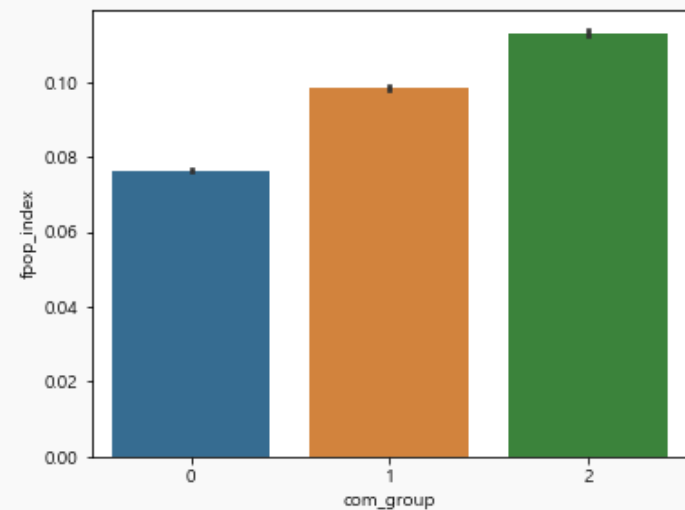
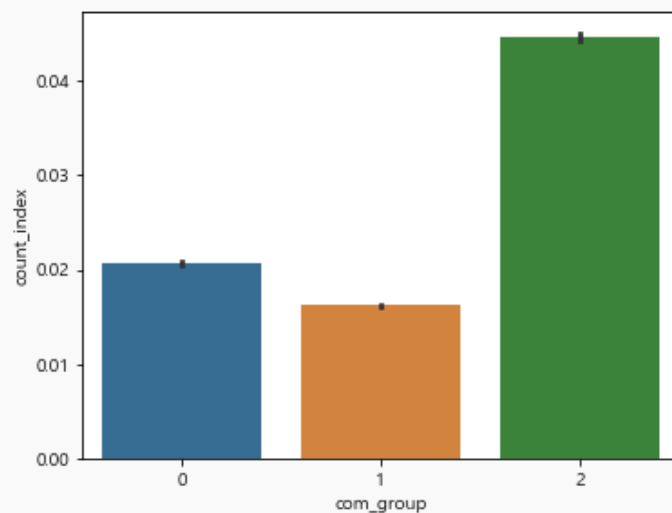
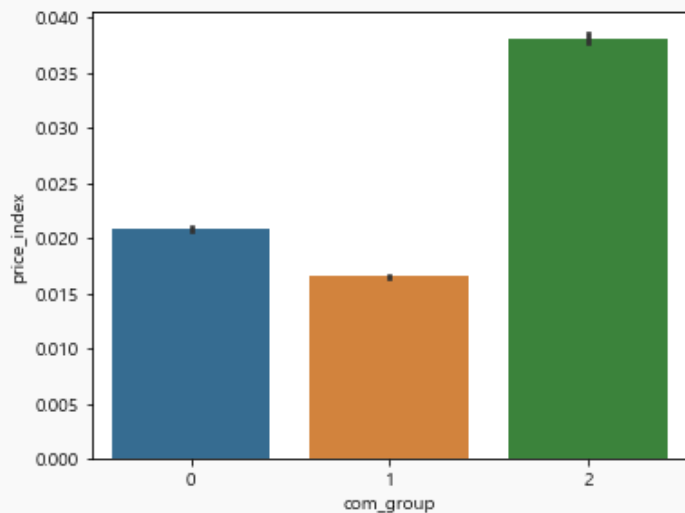
매출 관련 변수에 대해 군집된
그룹별 특성 파악



2그룹이 가장 높은 매출을 보임



데이터 분석 및 시각화



✓ 군집분석 된 그룹별로 매출 금액 지수, 매출 건수 지수, 유동인구 지수, 직장 인구 지수 확인

✓ 1그룹의 경우 유동인구는 많지만 전반적으로 매출이 낮게 도출

✓ 2그룹의 유동인구와 직장 인구가 가장 많은 것으로 보아 2그룹은 '핫플레이스'인 상권들의 그룹일 것이라고 예측 가능

데이터 분석 및 시각화

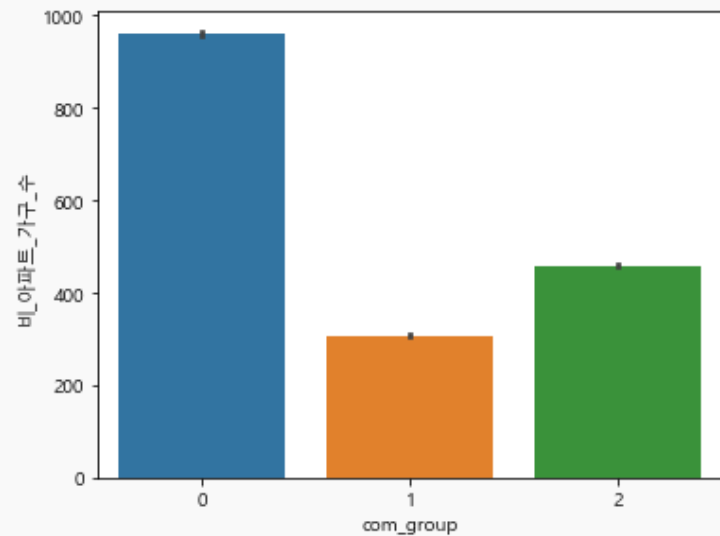
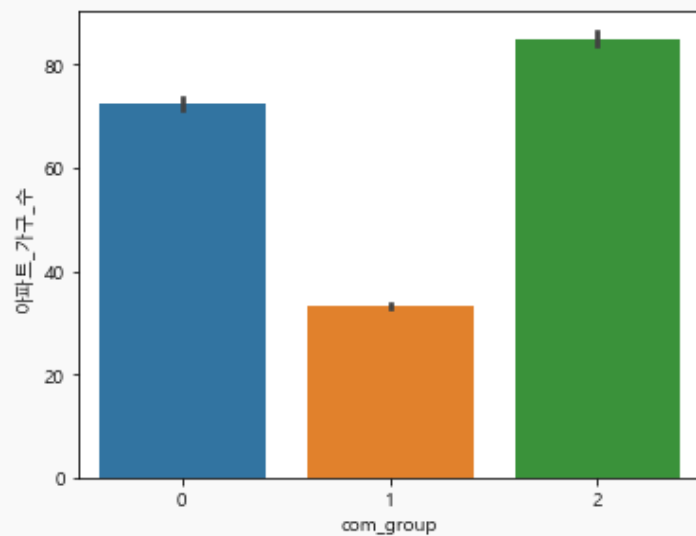
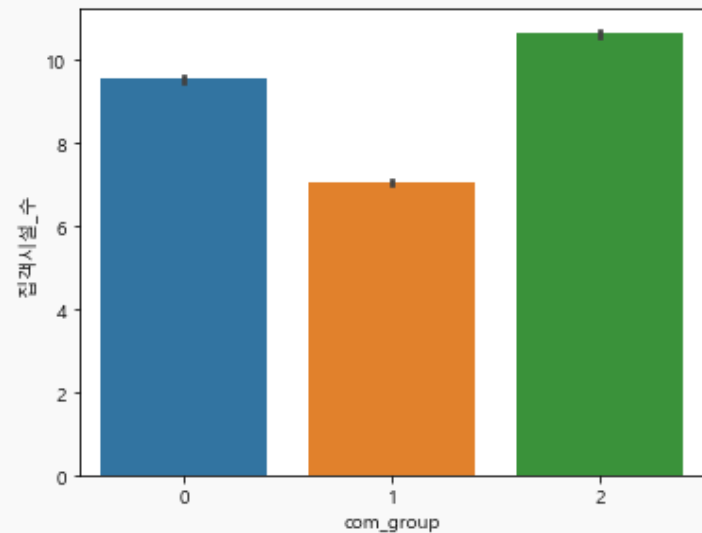
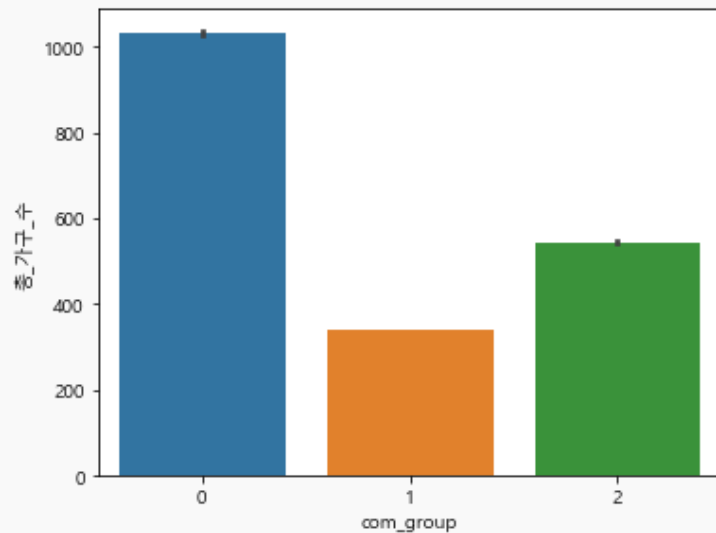
:군집분석



0그룹은 가구 수가 확연히 많고 집객시설 수도 많은 것으로 보아 상주인구가 많은 주거지역일 것으로 예측



2그룹은 가구 수에 비해 집객시설 수가 많은 것으로 보아 번화가일 가능성 多



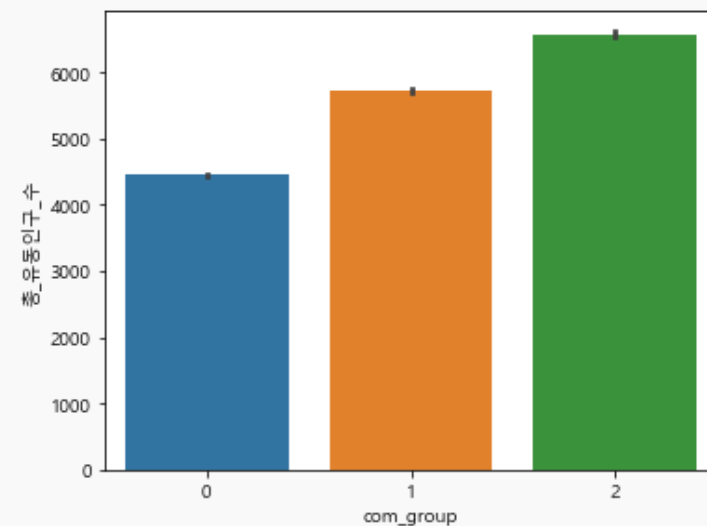
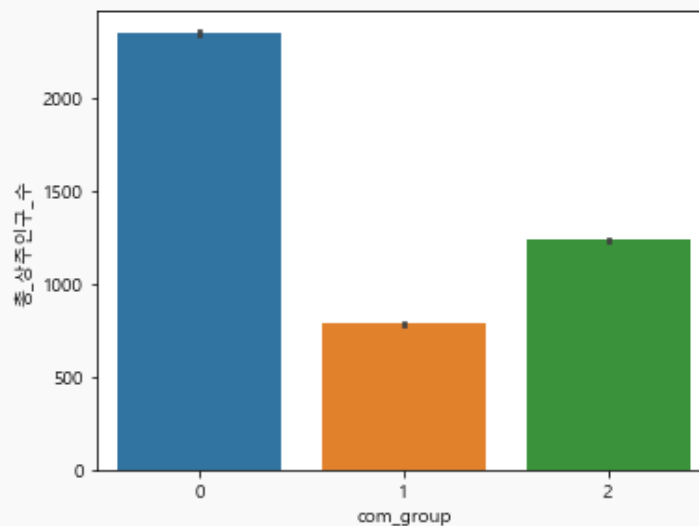
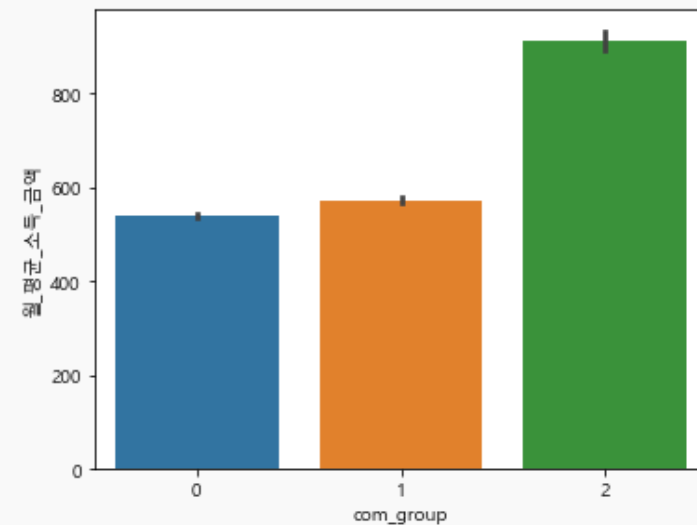
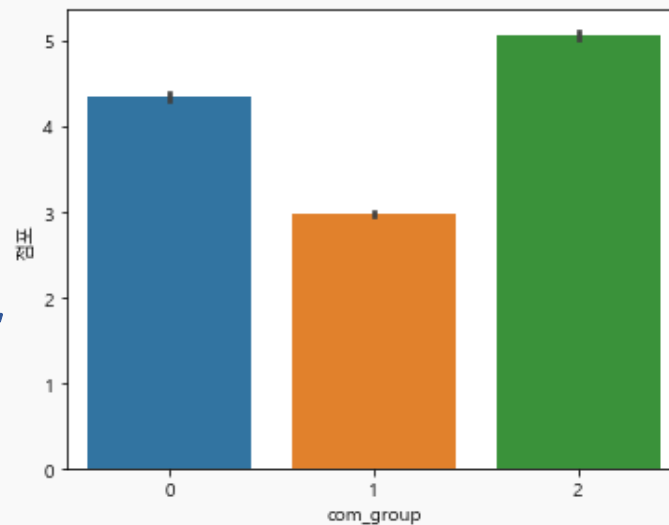
데이터 분석 및 시각화

:군집분석

✓ 군집분석된 그룹별로 점포 수, 월 평균 소득 금액, 총 상주인구 수, 총 유동인구 수 확인

✓ 총 상주인구 수는 총 가구 수와 같은 형태로 도출

✓ 2그룹은 상주인구가 많지는 않으나 월 평균 소득 금액이 높고 유동인구도 많은 편



분석결과 및 해석



분석 배경 및
목적

데이터 분석

분석결과 및
해석

참고자료

분석결과 및 해석

01

2년간 각 업종별 점포 수의 변화는 전반적으로 미미함

02

매출 금액과 건수에 따른 상권 또는 업종의 순위가 반드시 일치하지는 않음

03

2년간 매출 금액과 건수가 가장 높은 상권에서는 양식집의 비율이 가장 높게 도출

04

상권 위치 특성의 유사성에도 불구하고, 업종에 따라 매출 금액과 건수에 따른 순위 차이가 큼

05

점포별 매출 금액과 건수가 가장 높은 상권에서는 한식음식점의 비율이 가장 높게 나타남

분석결과 및 해석 : 군집분석

0그룹

- 점포별 평균 매출 가장 낮은 집단
- 총 가구 수와 비 아파트 가구 수 가장 높음
- 월 평균 소득 금액 가장 낮음

1그룹

- 총 매출 금액과 건수는 가장 낮으나, 점포별 평균 매출은 두 번째로 낮은 집단
- 직장인구, 상주인구, 가구 수, 집객시설, 아파트, 비 아파트 가장 낮음
- 그 외의 변수는 두 번째로 낮음

2그룹

- 총 매출과 점포별 평균 매출 모두 가장 높은 집단
- 유동인구와 직장인구, 집객시설 가장 높음
- 상주인구와 비 아파트 상대적으로 낮음

활성화 전략 제안

01

상권이 위치한 지역의 특성에 알맞은 업종의 분포가 상권 활성화에 중요하므로,
창업시 지역적 특성 고려한 업종 선택이 필요함

02

매출 지수가 높은 상권은 직장인구 수 역시 높게 나타나므로,
직장인들을 끌어들이 수 있는 마케팅 전략이 요구됨

03

매출 지수가 높은 상권은 집객시설 수 또한 높게 나타나므로,
상권의 활성화를 위해 관련 집객시설 증가를 위한 정책적 제안이 요구됨

04

소득수준은 상권의 매출지수로 귀결되므로,
저소득 계층을 위한 다양한 지원정책을 통해 골목상권의 매출 상승 효과를 기대할 수 있을 것

분석 한계점

01

다수의 이상치와 결측치로 데이터 분석에 제한

02

시간의 흐름에 따른 상권 변화를 반영하지 못한 채 분석 진행

03

이상치 제거 과정에서 일부 데이터 값 삭제

04

테이블 병합 과정에서 일부 데이터 중복 발생

참고자료



분석 배경 및
목적



데이터 분석



분석결과 및
해석



참고자료

참고자료

- 서울시 골목상권 매출액에 영향을 미치는 요인에 관한 연구(김현철, 이승일, 2019)
 - 골목상권의 매출액과 성장 결정요인 = Determinants of Sales and Growth of Alley Markets(강현모, 2018)
 - 발달상권과 골목상권에 위치한 음식점 생존과 폐업 비교 분석 : 이태원 지역을 중심으로(정동규, 2017)
 - 골목상권 영역 및 프로파일링 데이터 SET 파일 설명서(서울특별시 빅데이터 캠퍼스)
 - 서울시 골목상권 매출액에 영향을 미치는 요인에 관한 연구(김현철, 이승일, 2019)
 - 금융변수의 공통요인을 이용한 금융상황지수 개발(통화재정팀 차장 천병철, 조사역 박나연, 2007)
-