

도시재생의 전조 : 쇠퇴진단지표를 통한 예방적 접근

돌하르방분석단



Contents



1 서론

- 1) 도시재생사업
- 2) 기존 쇠퇴도 분석과의 차이
- 3) 도시재생의 전조



2 쇠퇴진단지표

- 1) 쇠퇴진단지표
- 2) 쇠퇴진단지표를 위한 시각화



3 선행지표

- 1) 선행지표의 필요성
- 2) 선행지표를 위한 인과성 검사
- 3) 주성분분석(PCA)과 클러스터링



4 결론

- 1) 분석 결과
- 2) 도시재생 사업 추진 제언





서

론

1 도시재생사업

도시재생의 정의

인구의 감소, 산업구조의 변화, 도시의 무분별한 확장, 주거환경의 노후화 등으로 쇠퇴하는 도시를 대상으로 지역 역량의 강화, 새로운 기능의 도입 및 창출, 지역 자원의 활용을 통해 경제적·사회적·물리적·환경적으로 활성화시키는 것



국가도시재생기본방침

- 국가도시재생 기본방침은 도시재생을 종합적·계획적·효율적으로 추진하기 위하여 국토교통부 장관이 수립하는 국가도시 재생전략
- 중점 시책: 도시 정책 패러다임 전환, 재정지원 확대, 금융지원·규제 완화, 역량 강화 및 공동체 활성화

2 기존 쇠퇴도 분석과의 차이

제주도의 지형적 특성



전통적인 해안도시



관광개발에 의한 성장

전통적인 해안도시이면서 관광개발이라는 사회적인 변수에 의한 도시 성장



지형적 장애요인
(화산지형과 오름 발달)



인위적 장애요인
(제주공항)

화산지형(오름의 발달 등)으로 인한 지형적 장애요인과
제주공항이라는 인위적 장애요인으로 인해 T자형 개발구조 형성

기존 쇠퇴도 분석과의 차이

한국도시지리학회지의 2007년 논문에 따르면
전통적인 취락구조는 내륙 쪽 보다는 용천대가 있는 해안을 따라 발달되었기 때문에
제주도는 해안지역을 중심으로 인구의 집중과 도시기능의 집적이 이루어진다.
그러나 기존 다수의 연구는 주로 내륙도시 위주였으므로
내륙도시의 모델을 해안도시에 적용하기에는 무리가 있다.

▲ 강민정(2007), " 제주시 도시화의 공간적 특성: 인구와 지가 변화를 중심으로", 한국도시지리학회지 10권 3월호, p.55-p.67

따라서, 기존 쇠퇴진단지표와 더불어
지역적 특성을 반영한 쇠퇴진단지표가 필요하다고 판단하였음

3 도시재생의 전조

선행지표의 필요성

미국의 도시쇠퇴 확산현상의 예처럼 도시쇠퇴 현상은 한 지역에서만 발생하는 현상이 아닌 공간적으로 확산 가능한 현상이며

국내에서도 이와 같은 도시쇠퇴 확산현상이 발생할 우려가 있다.

즉, 도시쇠퇴 현상은 공간적 차원의 현상이므로 미시적 요인 뿐 아니라 인근 지역을 포함하는 지역 차원의 요인들도 함께 고려될 필요가 있다.

▲ 엄현태, 우명제 "도시쇠퇴의 공간적 확산현상과 쇠퇴확산 요인에 대한실증분석" 국토계획 51.2 p.5-18 (2016) : 5.

- 쇠퇴진단지표의 선정은 쇠퇴한 지역을 판별하고 해당 지역의 재생을 목표로 설계됨
- 하지만 쇠퇴현상은 해당 지역의 쇠퇴 뿐 아니라 **인근 지역으로 확산** 될 수 있음
- 그렇다면, 선정된 쇠퇴진단지표에 대한 선행 지표들을 찾아서

쇠퇴가 가속화되기 이전에 먼저 재생사업이 필요한 지역을 구분하고 현재 쇠퇴지역과 함께 재생사업을 구상할 수 있다면,

지역 간 불균형적인 발전을 막고 도시쇠퇴의 확산을 감소시킬 수 있다고 판단함

3 도시재생의 전조

선행지표탐색

선행 지표는 아래의 세 가지 방법을 조합하여 분석하였음

① 인과성 검사

선정한 쇠퇴진단지표에 대한 **인과성 검사**를 진행하여 해당 쇠퇴진단지표의 원인이 되는 선행지표 발굴

② 주성분 분석(PCA)

선정한 쇠퇴진단지표를 독립변수로 갖는 집계구에 대하여 **주성분 분석(PCA)** 진행

③ 클러스터링

수행한 주성분 분석을 토대로 **클러스터링**을 통해 지역을 구분하고 쇠퇴진단지표 보조



쇠퇴진단지표

1 쇠퇴진단지표

기존 쇠퇴진단지표



인구 변화율



사업체수 변화율



노후 건축물 변화율

추가 쇠퇴진단지표



평일 유동인구 변화율



주말 유동인구 변화율



관광 사업체수 변화율

1 쇠퇴진단지표

추가 쇠퇴진단지표의 근거



평일유동인구 변화율

도시 쇠퇴지역을 선별하는 일은 반대적 현상인 도시활성화지역을 찾아내서 비교해 볼 수 있는 '유동인구'를 살펴야 한다. 고정적으로 거주하고 있는 인구와는 다르게 유동인구는 지역의 상황을 정확하게 판단할 수 있는 기준이 될 수 있다.

▲ 이민석(전남대학교 건축학부 교수), "한국형 도시재생, 도시재생 뉴딜사업이란", 공간정보 magazine vol.20



주말유동인구 변화율

"제품수명주기단계 변화율"을 참고하여 연간 관광객 수 변화율을 기준으로 관광지 발전 단계를 구분 관광객 수에 대한 집계구별 변화량 데이터를 구하기 어려워 관광객 수를 완벽히 대변할 수 없지만, 주말 유동인구의 변화는 관광 활동을 간접적으로 나타낼 수 있다고 판단하여 주말 유동인구 수로 대체

▲ 「이순자, 안소현, 유현아, 심창섭, 정광민(2021), "지역 관광지 쇠퇴원인 진단 및 활성화 방안 연구", 국토연구원」



관광사업체수 변화율

제주시는 해안도시라는 특성과 함께 관광개발, 도시계획에 의해 자연적인 성장이 아닌 계획적인 도시성장이 이루어지고 있으며, 이러한 자연적·인문적 요인을 반영하여 도시화가 진행되고 있다.

▲ 강민정(2007), "제주시 도시화의 공간적 특성: 인구와 지가 변화를 중심으로", 한국도시지리학회지 10권 3월호, p.55-p.67

1 쇠퇴진단지표

기존 쇠퇴진단지표



인구 변화율



사업체수변화율



노후건축물변화율

최근 5년간 3년 이상 연속으로 인구가 감소한 지역을 기준으로 하여
해당하는 집계구에 쇠퇴진단점수를 1점씩 부여

최근 5년간 3년 이상 연속으로 사업체 수가 감소한 지역이 거의 없기 때문에
집계구별 사업체 수 변화율을 (0,1)로 정규화 한 값을 반전시켜 쇠퇴진단점수로 부여

준공된 후 20년 이상 지난 건축물이 차지하는 비율을 기준으로 하여
해당하는 집계구에 쇠퇴진단점수로 부여

1 쇠퇴진단지표

추가 쇠퇴진단지표



평일유동인구변화율



주말유동인구변화율



관광사업체수변화율

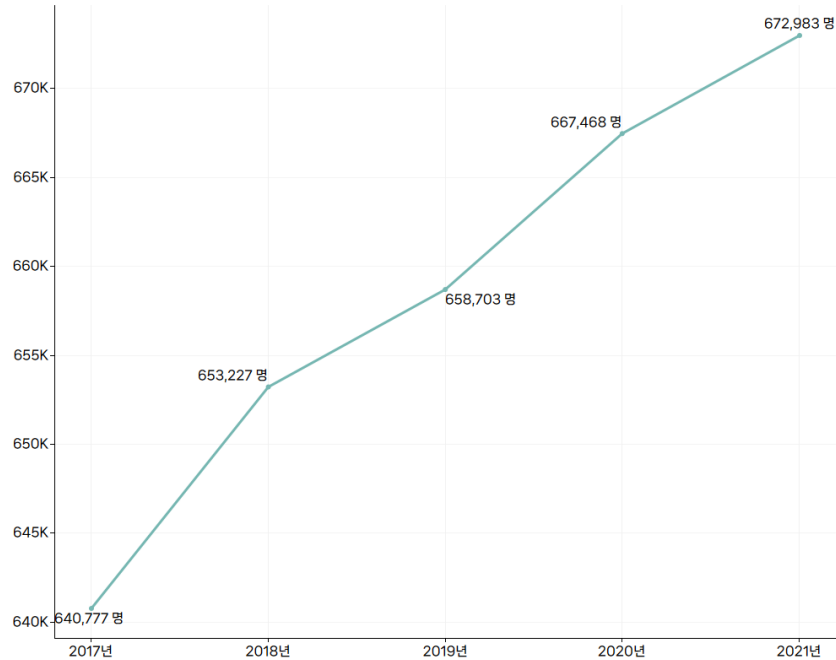
가장 최근 연도 기준의 평일 유동인구 변화율을 (0,1)로 정규화 한 값을 반전시켜 쇠퇴진단점수로 부여

가장 최근 연도 기준의 주말 유동인구 변화율을 (0,1)로 정규화 한 값을 반전시켜 쇠퇴진단점수로 부여

"관광도시"라는 제주도의 특성을 반영하기 위해 제주도 개·폐업 이력 데이터를 활용하여 관광 관련 업종만 필터링하여 가장 최근 연도의 관광사업체 수 변화를 (0,1)로 정규화 후 반전시켜 쇠퇴진단점수로 부여

2 쇠퇴진단지표를 위한 시각화

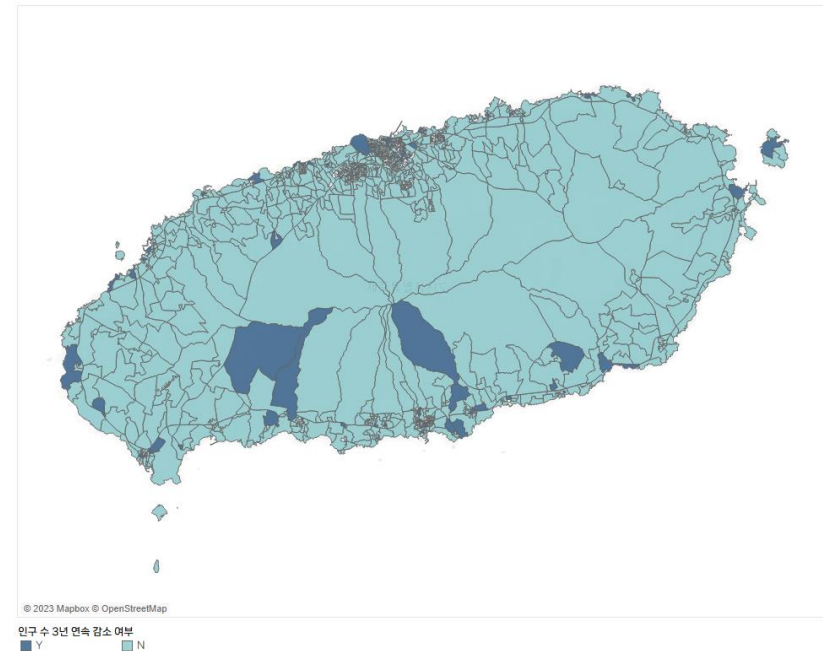
인구 수 시계열 그래프



- 인구 수 변화율 지표는 환경·경제·사회적 상황을 파악하는데 중요한 역할
- 쇠퇴하는 도시에서 공통적으로 인구 감소 현상이 나타남
⇒ 기본 쇠퇴진단지표 중 단연 **핵심 지표**
- 제주도 이주 트렌드 등의 이유로
총 인구 수가 꾸준히 증가하고 있는 것을 확인할 수 있음

① 인구 변화율

인구 수 3년 연속 감소 집계구

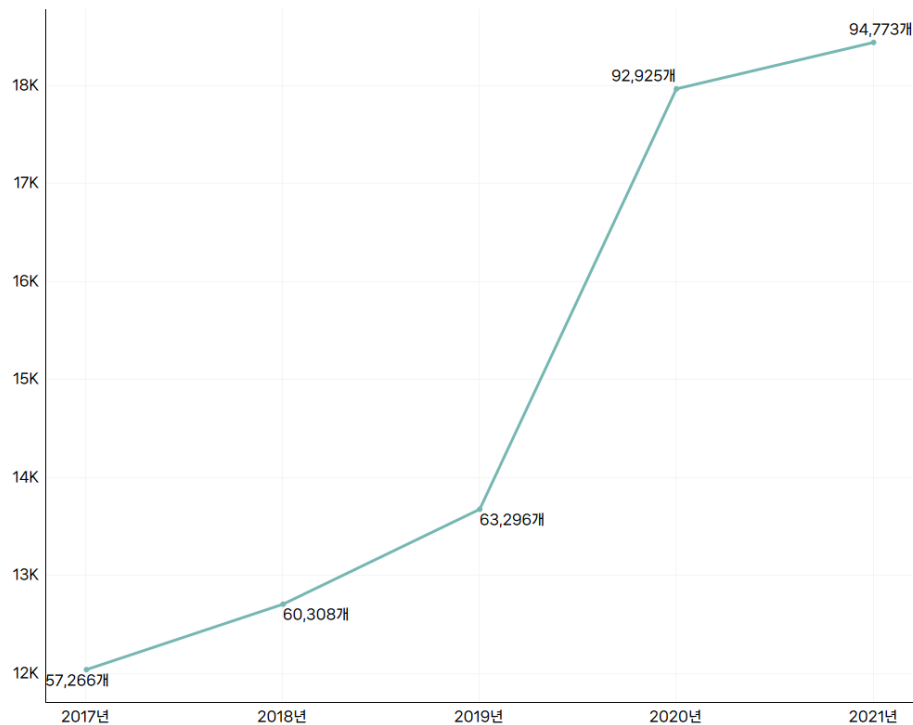


- 진한 파란 색으로 칠해진 집계구는 3년 연속으로
인구 수가 감소한 지역
- 3년 연속 감소한 집계구에 쇠퇴진단지표를 1점씩
부여하여 종합 점수에 반영

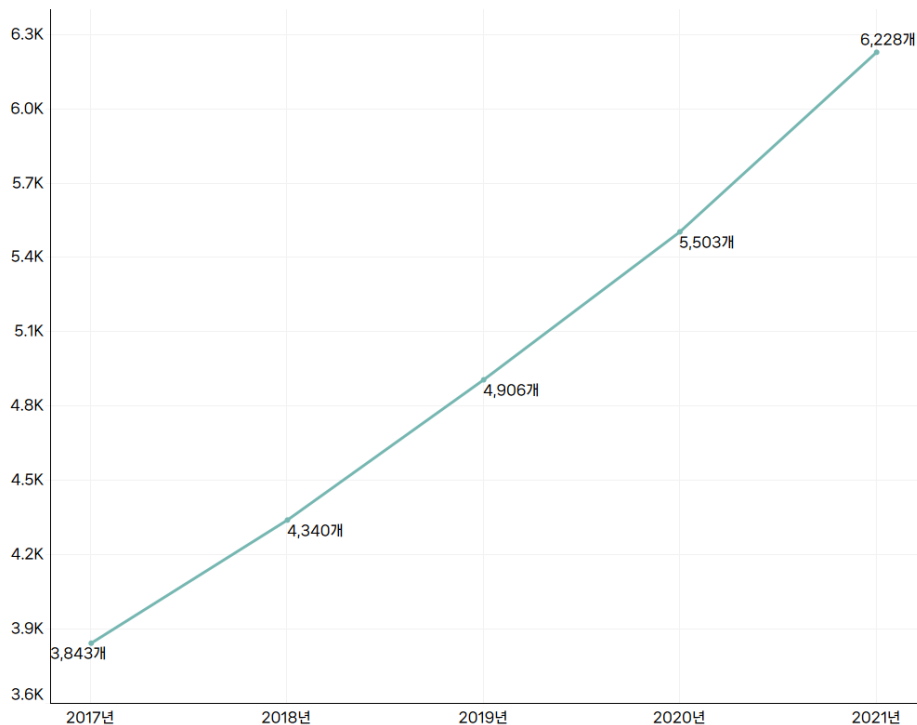
2 쇠퇴진단지표를 위한 시각화

② 사업체 수 변화율

사업체 수 시계열 그래프



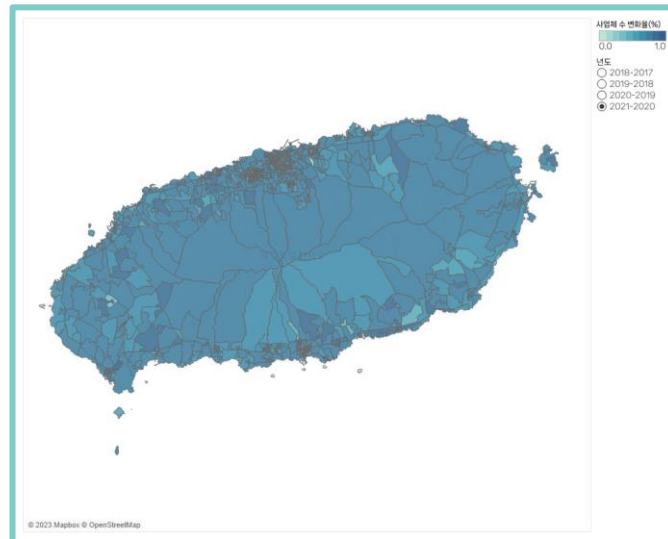
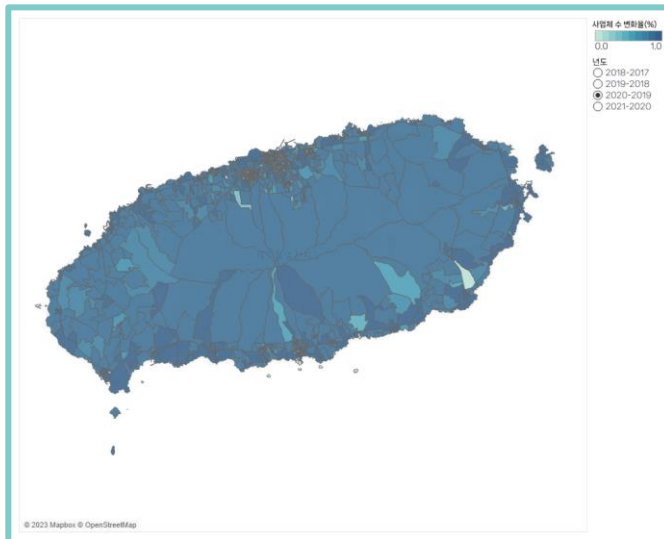
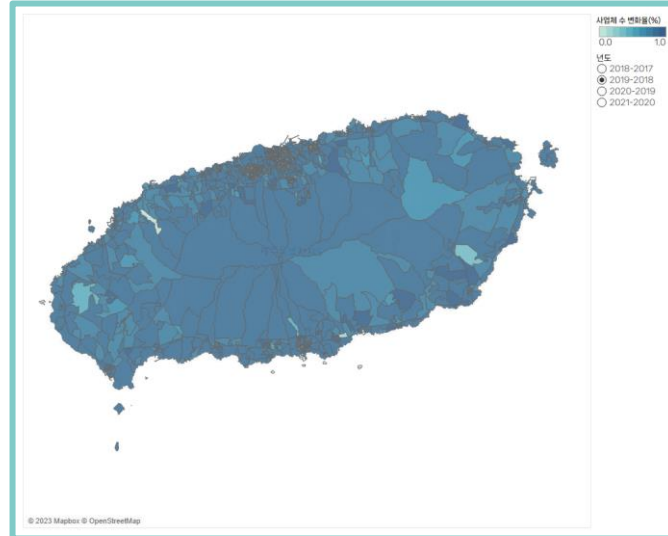
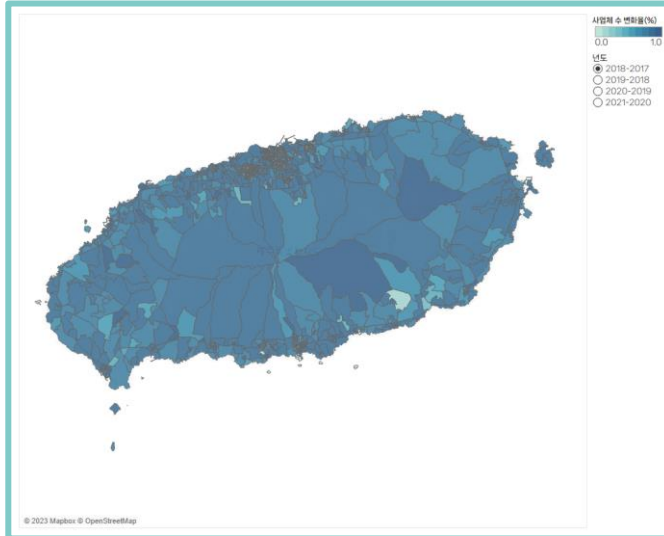
관광 사업체 수 시계열 그래프



- "도시 쇠퇴지수 개발에 대한 연구" 논문에 따르면 우리나라 중소도시의 주요 쇠퇴 원인은 도시 외곽개발로 인한 구도심 상권의 쇠락이라고 볼 수 있음
- 사업체 수 변화율 지표로 일반적인 "사업체 수 변화율"과
제주도라는 관광도시의 특성을 반영한 추가 쇠퇴진단지표인 "관광 사업체 수 변화율" 두 가지로 선정하였음

2 쇠퇴진단지표를 위한 시각화

연도별 사업체 수 변화율

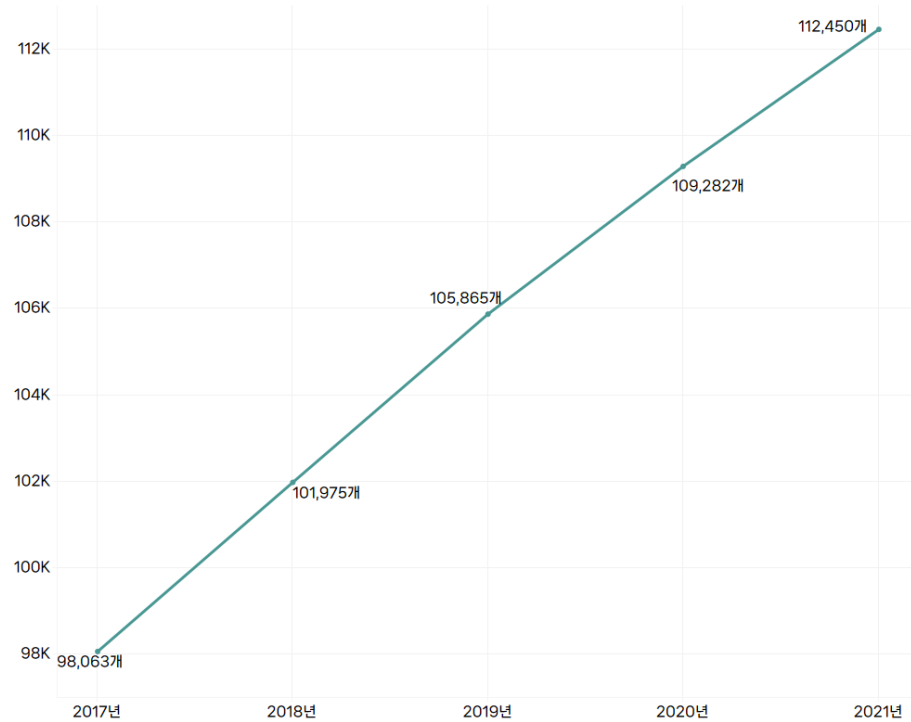


② 사업체 수 변화율

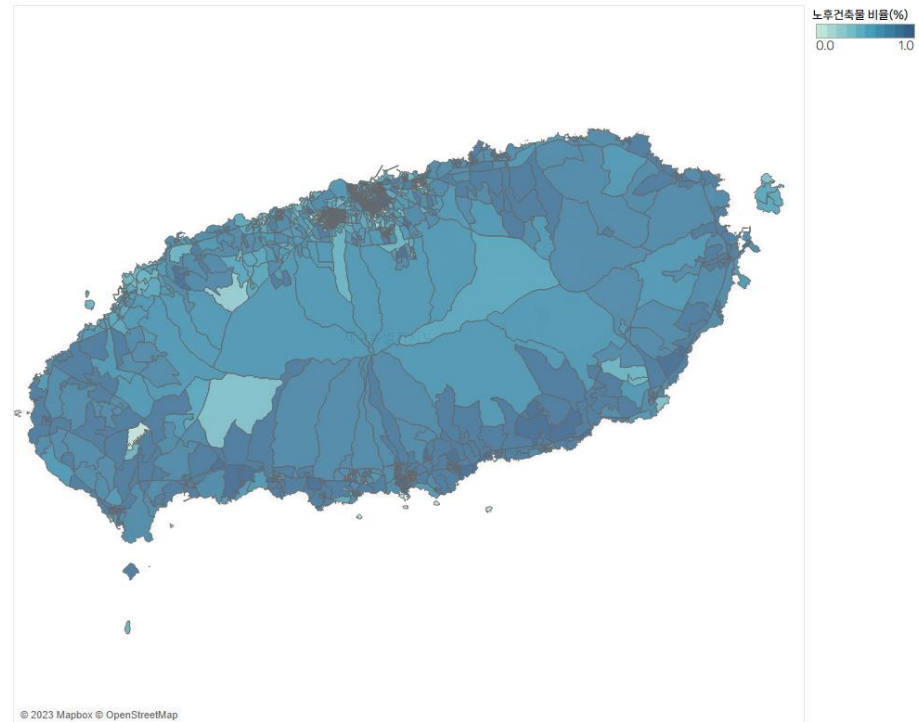
- 선행지표 파악을 위한 인과성 검사를 진행하기 위해 연도별로 사업체 수가 변화한 모습을 시각화 함
- 전년도 대비 사업체 수의 비율이 적은 집계구 순으로 sorting 하여 정규화 후 점수 부여

2 쇠퇴진단지표를 위한 시각화

노후 건축물 시계열 그래프



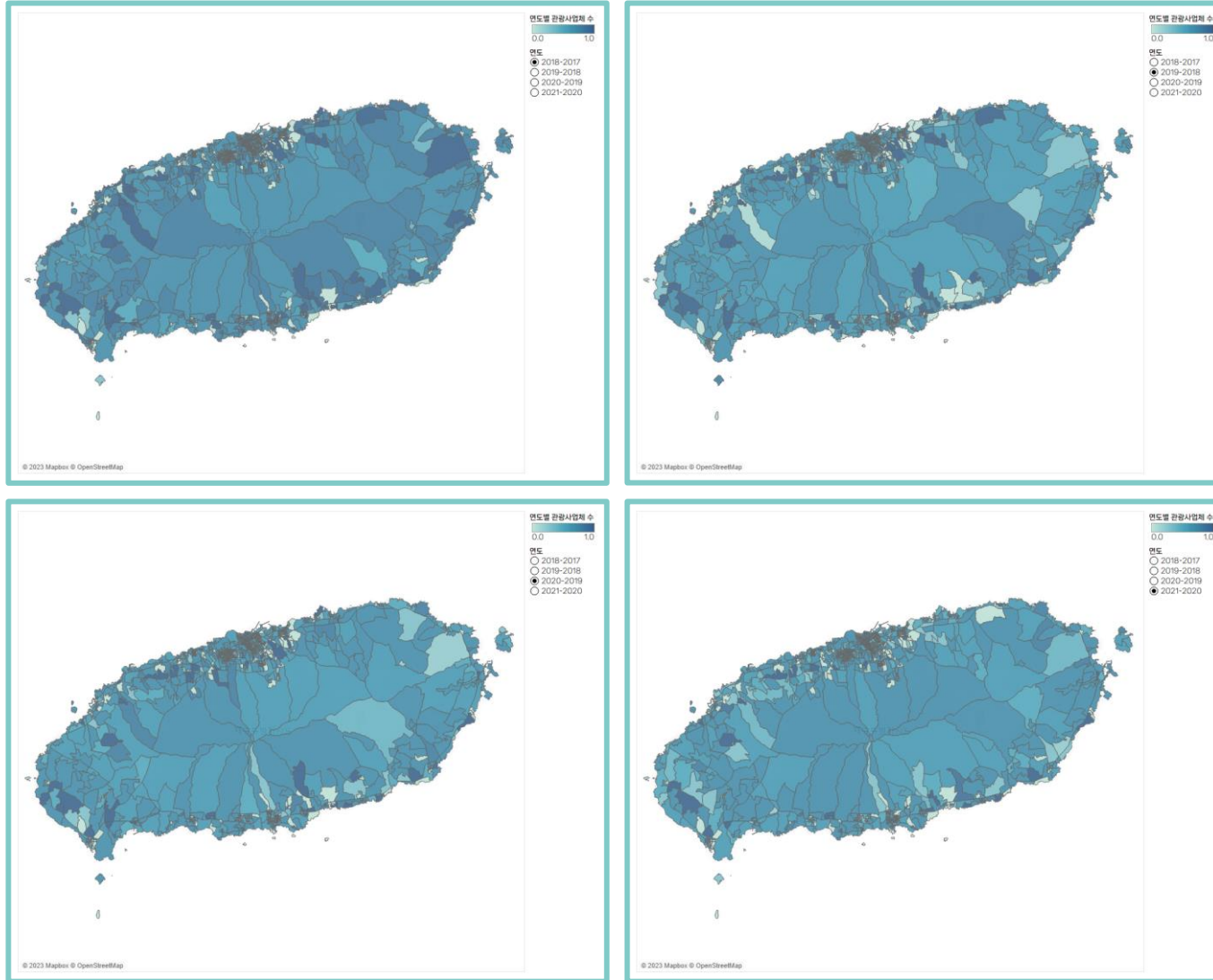
집계구별 노후 건축물 비율



- 당연히게도 시간이 지남에 따라 노후 건축물의 수는 증가하고 있는 추세
- 건물 노후도 데이터와 도로명주소(건물) 데이터를 집계구ID와 Geometry 데이터를 key로 하여 병합하였더니 중복 데이터가 거의 존재하지 않아 두 개의 데이터셋을 결합한 후 노후 건축물 비율을 구하여 해당 비율을 쇠퇴진단지표 점수로 반영함

2 쇠퇴진단지표를 위한 시각화

연도별 관광사업체 수 변화율

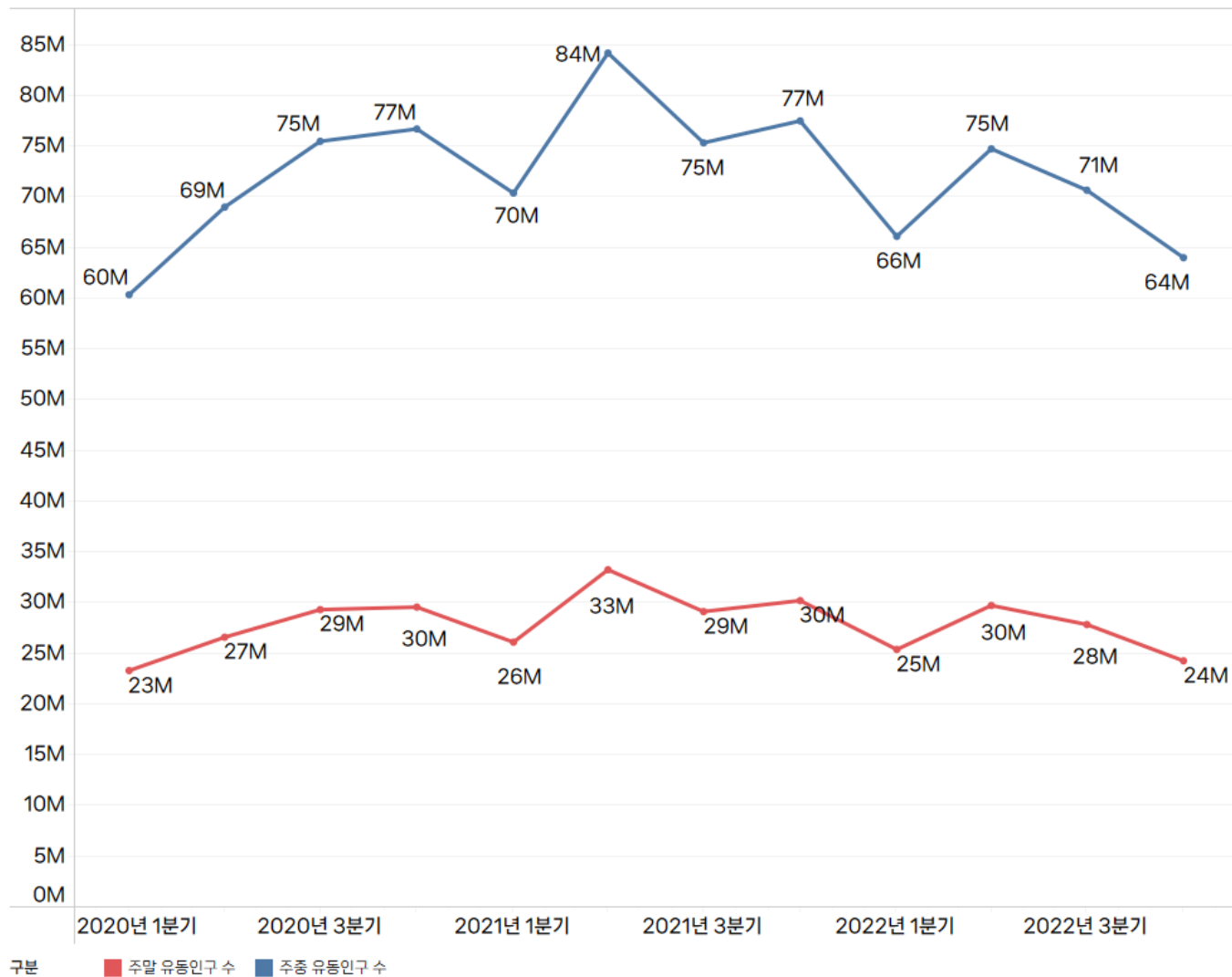


④ 관광사업체 수 변화율

- 사업체 수와 유사하게 전년도 대비 관광사업체 수의 비율이 적은 집계구 순으로 sorting 하여 정규화 후 점수 부여

2 쇠퇴진단지표를 위한 시각화

평일/주말 유동인구 시계열 그래프



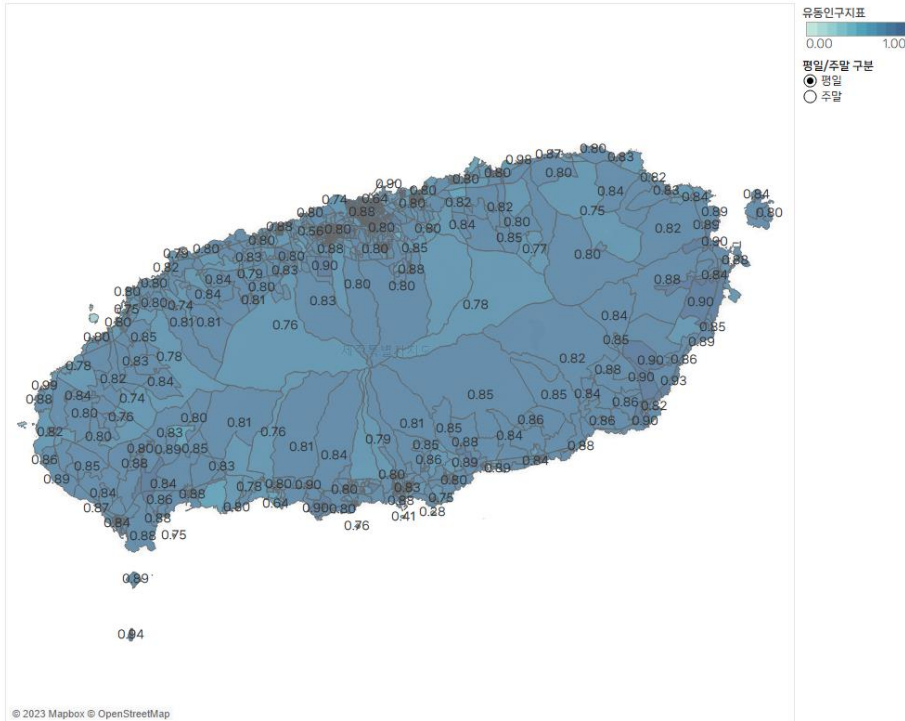
⑤ 유동인구 변화율

- 유동인구는 서론에서 설명한 바와 같이 환경·경제·사회 부분을 모두 대변할 수 있는 변수이므로 중요도가 높음
- 인구 수는 집계구의 환경적인 도시 재생 기능을 대변한다고 볼 수 있고 사업체 수의 변화는 경제적인 기능을 대변하지만, 단순히 인구 수와 사업체 수 만으로는 대중교통이나 도로의 발달과 같이 도시의 안정성과 편의성을 대변할 수는 없음

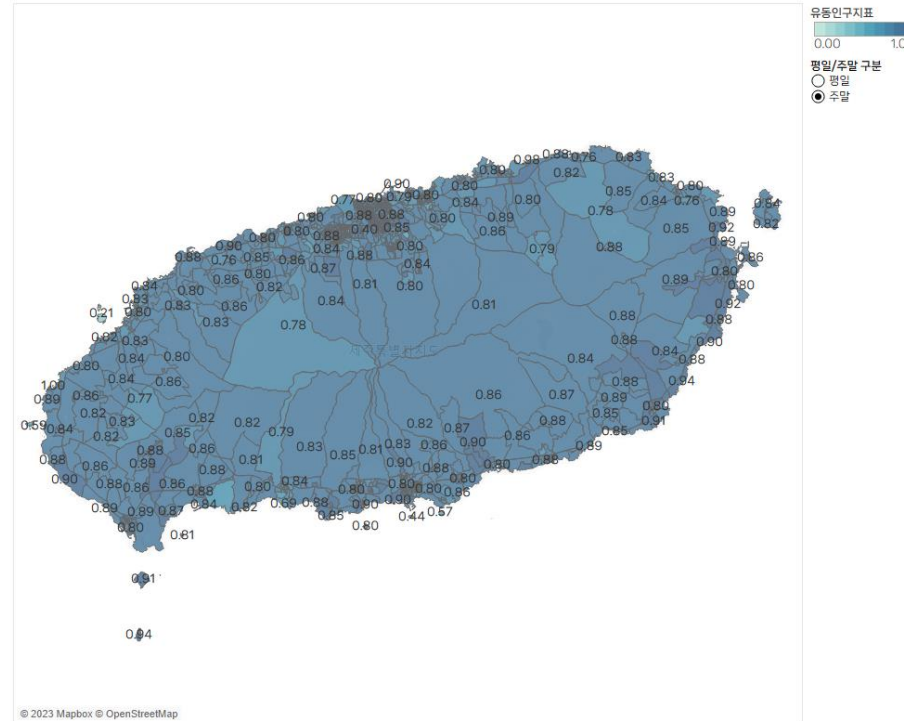
2 쇠퇴진단지표를 위한 시각화

⑤ 유동인구 변화율

평일 유동인구 변화율



주말 유동인구 변화율



- 이번 분석에서 핵심이라고 생각되는 “지역적 특성을 반영한 쇠퇴진단지표의 선정”을 위해 제주도가 가진 관광도시/해안도시의 특성을 고려하여 관광객 수 변화율이라는 지표를 선정함
- 「이순자, 안소현, 유현아, 심창섭, 정광민(2021), “지역 관광지 쇠퇴원인 진단 및 활성화 방안 연구”, 국토연구원」 논문의 “제품수명주기단계 변화율”을 참고하여 연간 관광객 수 변화율을 기준으로 관광지 발전 단계를 구분
- 관광객 수에 대한 집계구별 변화량 데이터를 구하기 어려워 관광객 수를 완벽하게 대변할 수 없지만 주말 유동인구의 변화는 관광 활동을 간접적으로 나타낼 수 있다고 판단하여 주말 유동인구 수로 대체하였음



선행 지표

1 선행 지표의 필요성

선행 지표

선행 지표(Leading Indicator)란 경제·과학·기술·사회 등 다양한 분야에서 미래의 특정 사건이나 현상을 예측하기 위해 사용하는 지표를 말함
이러한 선행 지표는 일반적으로 사건이나 현상이 발생하기 전에 변동하는 경향이 있어, 미래를 예측하는 데 도움을 줌
하지만 이 지표가 항상 정확한 예측을 하는 것은 아니기 때문에, 다른 통계적 방법과 조합하는 것이 필요하므로 클러스터링 기법을 함께 활용함

선행 지표의 필요성

따라서, 앞서 선정된 쇠퇴진단지표의 선행 지표를 구축하여 쇠퇴 가능성 예측과 조기 경고 시스템을 구축하는 데 도움을 줄 수 있는 전략을 수립할 수 있음
원도심이 구도심으로 변해가는 과정을 예측함으로써, 두 지역의 균형적 발전을 도모할 수 있음

선행 지표 구축 대상

기본 쇠퇴진단지표			추가 쇠퇴진단지표		
인구 변화율	사업체 수 변화율	노후 건축물 수 변화율	평일 유동인구 변화율	주말 유동인구 변화율	관광 사업체 수 변화율

2 선행 지표를 위한 인과성 검사

인과성 검사

인과 관계는 변수 간 상관 관계를 충족하면서 원인과 결과의 관계 파악이 가능하므로, 이를 토대로 **선행 지표를 예측**할 수 있다고 판단

Granger 인과관계 검정

시차분포모형(Distributed Lag Model)*을 이용하여 원인과 결과를 확인할 수 있는 검정 방법

- 시차분포모형이란, 회귀모형이 설명변수의 현재 관측 값은 물론, 과거의 관측 값까지 포함하고 있는 설명변수들로 정의된 경우의 모형

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_n x_{t-n} + \epsilon_t$$

과거의 사건은 현재의 사건을 발생시킬 수 있지만, 미래의 사건은 현재의 사건을 발생시킬 수 없다는 것을 근거로 하여
⇒ 변수 X가 변수 Y에 대하여 Granger 원인이라면, X의 과거 정보는 Y의 현재 값을 예측하는 데 도움을 줌

데이터를 기반으로 구한 쇠퇴진단지표 중 인과성에 대한 가설을 설정하여, Granger 인과성 검사 결과와 외부 논문을 기반으로 하여 결론을 도출하고자 함

Granger 인과성 검사 절차

가설 설정

데이터 준비

모델 설정

F-값 계산

p-value 계산

2 선행 지표를 위한 인과성 검사

Granger 인과성 검사 절차

1. 가설 설정

- ① 귀무가설(H_0): X가 Y에 영향을 미치지 않는다. 대립가설(H_1): X가 Y에 영향을 미친다.
- ② X가 Y에 대하여 Granger 인과관계가 성립하고, Y가 X에 대하여 Granger 인과관계가 성립하지 않는 경우를 말함

2. 데이터 준비

- ① pandas 패키지의 pivot-table을 활용하여 컬럼을 같은 범위의 연도로, 인덱스를 집계구로, 값을 카운트로 하여 같은 형식으로 변환
- ② X데이터와 Y데이터를 concat하여 데이터프레임을 결합

3. 모델 설정

- ① statsmodel 패키지의 grangercausalitytests를 이용하여 X와 Y의 인과성을 검사
- ② 보유한 데이터는 연도별로 되어 있어 데이터 양이 충분하지 않아 Granger 인과성 검사의 지연(lag)값을 1년으로 설정
⇒ 쇠퇴진단지표가 선행지표의 1년 뒤에 영향을 줄 수 있음을 의미

4. F-값 계산 및 p-value 계산

- ① 검정 통계량(F-값)과 유의확률(p-value)을 계산 ⇒ 유의수준 0.05로 설정

5. 결론 도출

- ① $p\text{-value} \leq \text{유의수준}(\alpha)$: 귀무가설 기각. 대립가설은 통계적으로 유의미하다고 간주
- ② $p\text{-value} \geq \text{유의수준}(\alpha)$: 귀무가설 기각하지 않음. 대립가설은 통계적으로 유의미하지 않다고 간주
- ③ X가 Y에 선행하고 Y가 X에 선행하지 않는 조건을 모두 만족하는 데이터에 한하여 Granger 인과관계가 있다고 판단

2 선행 지표를 위한 인과성 검사

Y: 인구 수

X: 유소년 인구 수

Testing if population Granger-causes youth:

Granger Causality

number of lags (no zero) 1

ssr based F test: F=0.7600 , p=0.5435 , df_denom=1, df_num=1

ssr based chi2 test: chi2=3.0399 , p=0.0812 , df=1

likelihood ratio test: chi2=2.2612 , p=0.1327 , df=1

parameter F test: F=0.7600 , p=0.5435 , df_denom=1, df_num=1

Testing if youth Granger-causes population:

Granger Causality

number of lags (no zero) 1

ssr based F test: F=504.9795, p=0.0283 , df_denom=1, df_num=1

ssr based chi2 test: chi2=2019.9180, p=0.0000 , df=1

likelihood ratio test: chi2=24.9060 , p=0.0000 , df=1

parameter F test: F=504.9795, p=0.0283 , df_denom=1, df_num=1

X: 생산가능 인구 수

Testing if population Granger-causes adult:

Granger Causality

number of lags (no zero) 1

ssr based F test: F=0.8491 , p=0.5260 , df_denom=1, df_num=1

ssr based chi2 test: chi2=3.3965 , p=0.0653 , df=1

likelihood ratio test: chi2=2.4588 , p=0.1169 , df=1

parameter F test: F=0.8491 , p=0.5260 , df_denom=1, df_num=1

Testing if adult Granger-causes population:

Granger Causality

number of lags (no zero) 1

ssr based F test: F=1108.2835, p=0.0191 , df_denom=1, df_num=1

ssr based chi2 test: chi2=4433.1341, p=0.0000 , df=1

likelihood ratio test: chi2=28.0459 , p=0.0000 , df=1

parameter F test: F=1108.2835, p=0.0191 , df_denom=1, df_num=1

X→Y	p-value	귀무가설 기각 여부	Y→X	p-value	귀무가설 기각 여부	결론
유소년 인구 수 → 인구 수	0.0283	O	인구 수 → 유소년 인구 수	0.5435	X	X는 Y에 선행함
생산가능 인구 수 → 인구 수	0.0191	O	인구 수 → 생산가능 인구 수	0.5260	X	X는 Y에 선행함

2 선행 지표를 위한 인과성 검사

Y: 사업체 수 사업체 수에 선행되는 유의미한 지표는 존재하지 않는 것으로 보임

Y: 관광 사업체 수 주택 수와 생산가능 인구 수 모두 유의수준과 근사한 값을 가지지만 귀무가설을 기각하지 못함

X: 주택 수

Testing if tourCompany Granger-causes housing:

Granger Causality

number of lags (no zero) 1

ssr based F test: F=0.0001 , p=0.9927 , df_denom=1, df_num=1

ssr based chi2 test: chi2=0.0005 , p=0.9818 , df=1

likelihood ratio test: chi2=0.0005 , p=0.9818 , df=1

parameter F test: F=0.0001 , p=0.9927 , df_denom=1, df_num=1

Testing if housing Granger-causes tourCompany:

Granger Causality

number of lags (no zero) 1

ssr based F test: F=93.8933 , p=0.0655 , df_denom=1, df_num=1

ssr based chi2 test: chi2=375.5733, p=0.0000 , df=1

likelihood ratio test: chi2=18.2110 , p=0.0000 , df=1

parameter F test: F=93.8933 , p=0.0655 , df_denom=1, df_num=1

X: 생산가능 인구 수

Testing if tourCompany Granger-causes adult:

Granger Causality

number of lags (no zero) 1

ssr based F test: F=0.0488 , p=0.8616 , df_denom=1, df_num=1

ssr based chi2 test: chi2=0.1951 , p=0.6587 , df=1

likelihood ratio test: chi2=0.1905 , p=0.6625 , df=1

parameter F test: F=0.0488 , p=0.8616 , df_denom=1, df_num=1

Testing if adult Granger-causes tourCompany:

Granger Causality

number of lags (no zero) 1

ssr based F test: F=77.3002 , p=0.0721 , df_denom=1, df_num=1

ssr based chi2 test: chi2=309.2010, p=0.0000 , df=1

likelihood ratio test: chi2=17.4422 , p=0.0000 , df=1

parameter F test: F=77.3002 , p=0.0721 , df_denom=1, df_num=1

X→Y	p-value	귀무가설 기각 여부	Y→X	p-value	귀무가설 기각 여부	결론
주택 수 → 관광 사업체 수	0.0655	X	관광 사업체 수 → 주택 수	0.9927	X	X는 Y에 선행하지 않음
생산가능 인구 수 → 관광 사업체 수	0.0721	X	관광 사업체 수 → 생산가능 인구 수	0.8616	X	X는 Y에 선행하지 않음

2 선행 지표를 위한 인과성 검사

Granger 인과성 검사 결과

X→Y	p-value	귀무가설 기각 여부
생산가능 인구 수 → 인구 수	0.0191	O
유소년 인구 수 → 인구 수	0.028	O
주택 수 → 관광 사업체 수	0.066	X
생산가능 인구 수 → 관광 사업체 수	0.072	X

- 인구수를 제외한 선정된 다른 모든 쇠퇴진단지표에서 귀무가설을 기각한 결과는 얻지 못했으나, 유의미한 의미를 가진 결과도 상당 부분 존재함
- 물론 인과성 검사만으로는 선행지표를 채택할 수 없고
주성분 분석 및 클러스터링을 통해 추가적인 분석이 필요할 것으로 보임
- 위의 결과로 얻을 수 있는 통찰이 있다면
제주도의 도시재생사업은 노후 건물 보수와 같은 경제적인 형태보다
유소년 인구, 즉 아이를 키우기 좋은 환경을 조성하는 것이 중요하다는 점을 알 수 있음
- 인구 수 중 유소년 인구 수의 비중이 증가하는 도시는 앞으로 인구 수가 증가하고
도시 쇠퇴 점수를 낮출 수 있다고 생각함

3 주성분 분석(PCA)과 클러스터링

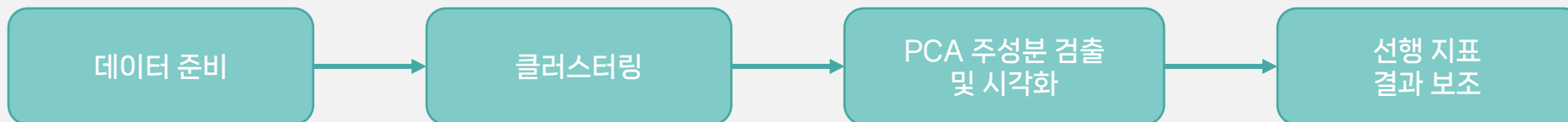
주성분 분석(PCA)

- 주성분 분석(Principal component Analysis; PCA): 직교 변환을 사용하여 고차원 데이터를 저차원 데이터로 환원시키는 기법

주성분 분석 이유

- 도시 쇠퇴는 복잡한 현상으로 다양한 요인이 상호 작용하여 발생함
- 이를 정확하게 이해하고 대응하기 위해서 다수의 변수(도시 쇠퇴진단지표)를 고려해야 함
- 다수의 변수를 고려하게 되면 분석이 복잡해지고 다중공선성 등 통계적 문제를 야기할 수 있는데, PCA는 이러한 문제를 해결하기 위한 효율적 접근 방법임
- 또한, PCA로 차원을 줄여 압축된 성분을 토대로 진행한 클러스터링을 시각화 하려고 함

분석 순서



3 주성분 분석(PCA)과 클러스터링

① 데이터 준비

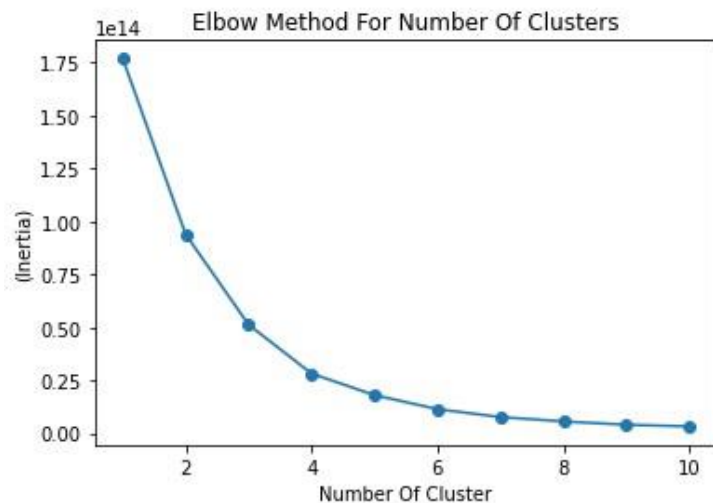
- 6개의 도시 쇠퇴진단지표 (인구 수, 사업체 수, 노후 건축물 수, 평일 유동인구 수, 주말 유동인구 수, 관광업체 수)를 독립변수로 하여 주성분을 추출하기 위한 데이터프레임 생성

② 클러스터링

- 쇠퇴진단지표로 인덱스가 집계구 ID이고 6개의 진단지표 변수를 독립변수로 한 데이터프레임을 생성하여 적절한 클러스터(클래스) 수를 정하기 위해서 Elbow Method*를 사용

※ Elbow Method: 군집분석(특히, Kmeans 분석)에서 각각의 클러스터 내 포인트들의 거리의 제곱합으로 군집 수를 결정하는 방법을 말함

- 클러스터 개수를 변화시켜가며 Elbow Method를 실행한 결과, 5개 이후로는 클러스터 수를 증가시켜도 클러스터 내 분산의 변화가 미미 했기 때문에 클러스터 개수를 5개로 확정함



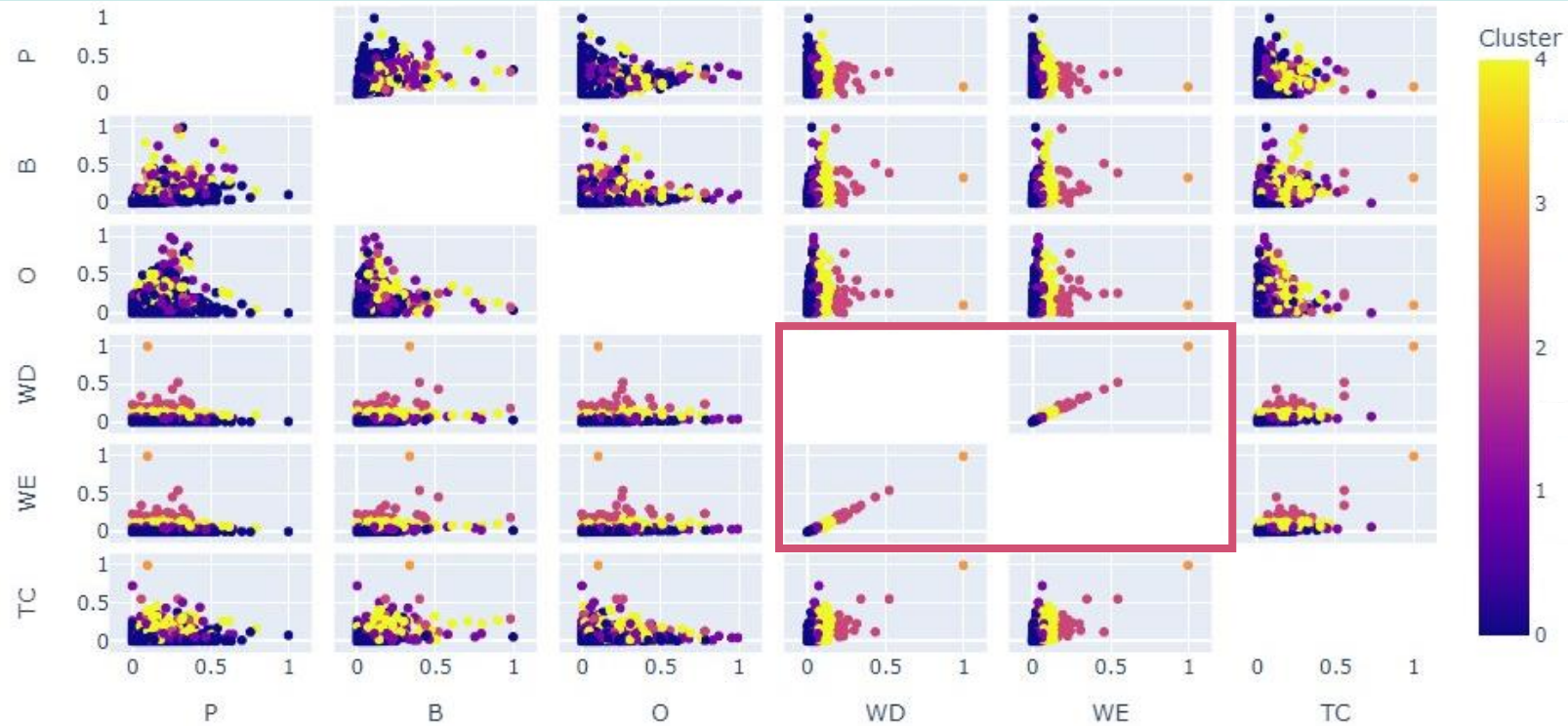
```
num_cluster=5 # 원하는 클러스터 수
kmeans=KMeans(n_clusters=num_clusters, random_state=42)
cdf2021['Cluster']=kmeans.fit_predict(cdf2021)
```

	population2021	business2021	oldbuilding2021	moving_weekDay2021	moving_weekEnd2021	tourCompany2021	Cluster
39010110010001	412.00	18.00	12.00	325111.10	119292.50	2.00	1
39010110010002	299.00	21.00	70.00	163569.00	59506.60	3.00	0
39010110020001	374.00	32.00	119.00	337890.20	150656.30	2.00	1
39010110020002	427.00	47.00	100.00	334860.10	130987.90	2.00	1
39010110020003	401.00	35.00	66.00	144955.60	57483.50	0.00	0
...
39020620020001	259.00	44.00	121.00	158488.80	71564.50	3.00	0
39020620020004	222.00	11.00	168.00	53589.50	20622.30	0.00	0
39020620020005	413.00	51.00	172.00	168131.60	67404.10	2.00	0
39020620020006	1083.00	118.00	459.00	496880.90	208100.00	15.00	4
39020620020008	587.00	94.00	247.00	439415.80	184916.50	9.00	1

3 주성분 분석(PCA)과 클러스터링

③ PCA 주성분 검출 및 시각화

도시쇠퇴진단지표 상관 관계 그래프



- 각 Pair Plot 을 보면 WD(평일 유동인구), WE(주말 유동인구)만 양의 상관관계를 가짐
- 즉, 유동인구의 상관관계를 통해 제주도가 관광도시의 특성을 가진 것을 확인 할 수 있음
- 또한 Pair Plot 을 통해 집계구를 구분한 클러스터링을 확인할 수 있음

3 주성분 분석(PCA)과 클러스터링

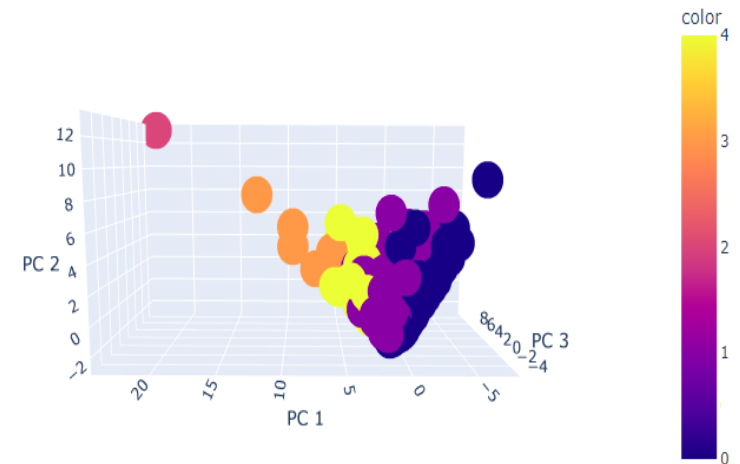
③ PCA 주성분 검출 및 시각화

- 1) PCA를 단계적으로 실행하여 공분산행렬을 통해 고유 값과 고유 벡터를 구할 수 있음
이 때 고유 값이 설명 가능한 분산량에 해당함
- 2) 따라서, $PCA=PCA(n_components=n)$, 즉 n 값을 늘려가며 분석을 진행
- 3) 만약 추가 주성분의 분산량이 극소량일 경우에 추가 주성분을 제외한 n 값으로 주성분 개수를 채택
 $n=4$ 이후부터 추가적인 주성분의 분산량은 무시할 만한 수치였으므로, **주성분을 4개로 채택**

주성분(PC) 1의 설명된 분산비율	0.37	주성분(PC) 1의 주요 성분	주말 유동인구 수
주성분(PC) 2의 설명된 분산비율	0.35	주성분(PC) 2의 주요 성분	인구수
주성분(PC) 3의 설명된 분산비율	0.13	주성분(PC) 3의 주요 성분	관광 업체 수
주성분(PC) 4의 설명된 분산비율	0.11	주성분(PC) 4의 주요 성분	사업체 수

- 4) 앞서 수행한 클러스터링과 주성분 분석을 바탕으로 시각화

Total Explained Variance: 97.12%



3 주성분 분석(PCA)과 클러스터링

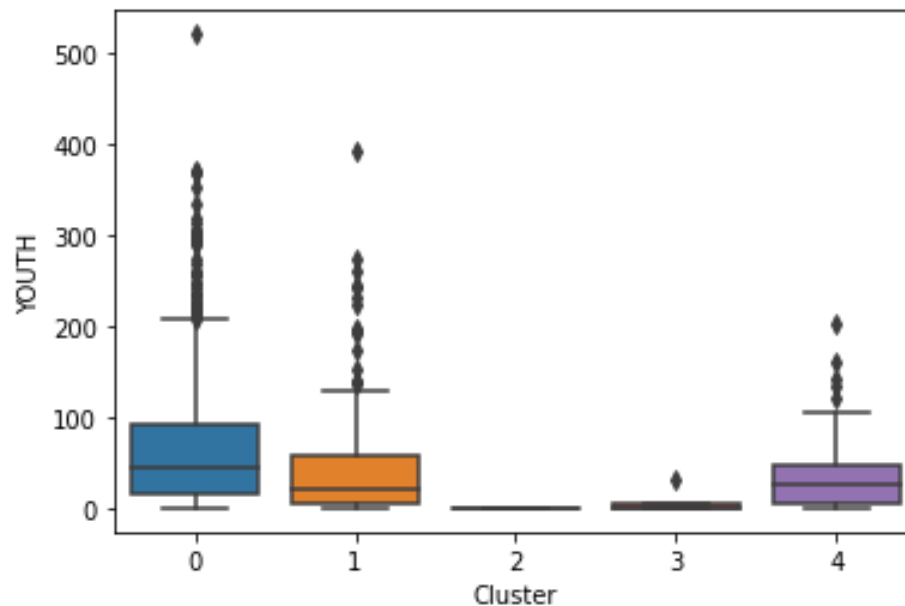
④ 선행 지표 결과 보조 클러스터링과 유소년 인구수의 관계

- 1) 우선 진단지표의 특성을 기반으로 클러스터링을 수행하여 지역을 여러 5개의 그룹으로 분류하였음
- 2) 그 다음 ANOVA 통계 검정을 통해 각 클러스터 간 '유소년 인구'의 평균 차이를 분석
- 3) F-통계량은 12.37로 클러스터 간 '유소년 인구'의 평균에 유의미한 차이가 있었고, p-value는 7.36e-10으로 통계적으로 유의미한 값을 가짐

```
clusters = foryouth['Cluster'].unique()
data_arrays = [foryouth[foryouth['Cluster'] == cluster]['YOUTH'] for cluster in clusters]

# ANOVA
f_stat, p_val = stats.f_oneway(*data_arrays)
print(f"F-statistic: {f_stat}, P-value: {p_val}")

F-statistic: 12.3657404489963, P-value: 7.360284087594768e-10
```

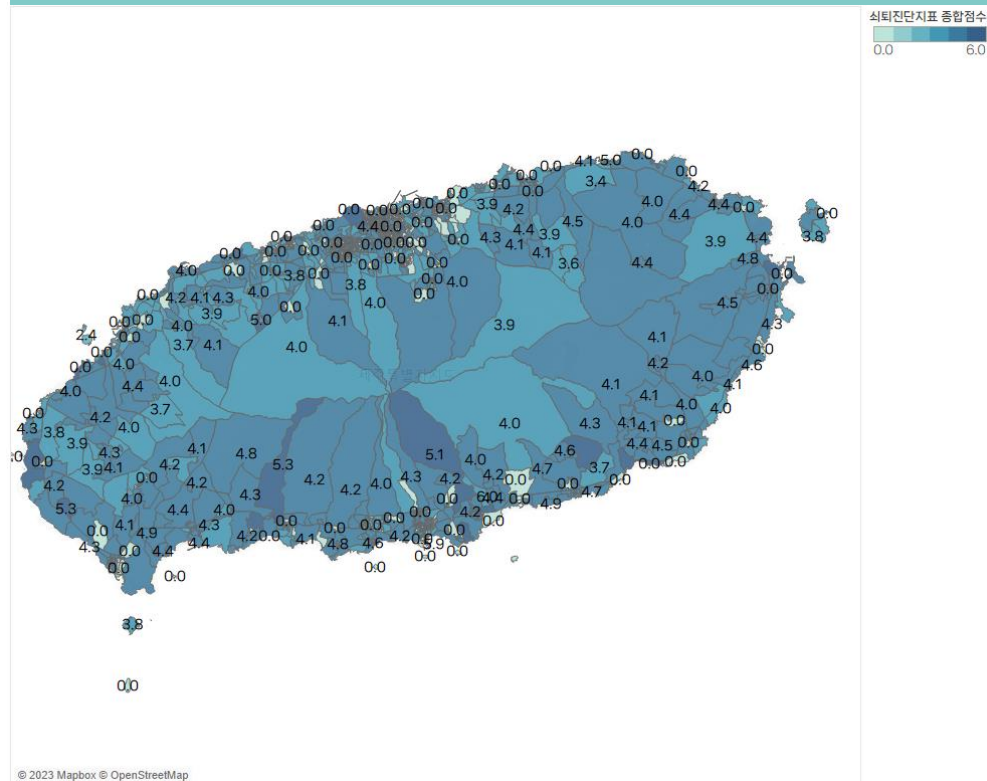


- 결론적으로 선행지표인 '유소년인구'는 지역을 그룹화 하는 데 중요한 특성을 가지고 있다고 볼 수 있음
 - 따라서 검정결과는 유소년인구가 선행지표로 활용되는데 뒷받침이 되는 근거가 될 수 있다고 보임
- 또한, 다른 쇠퇴점수는 상대적으로 높지만 유소년 인구 점수가 낮은 그룹에 대해서는 상세한 분석이 필요함

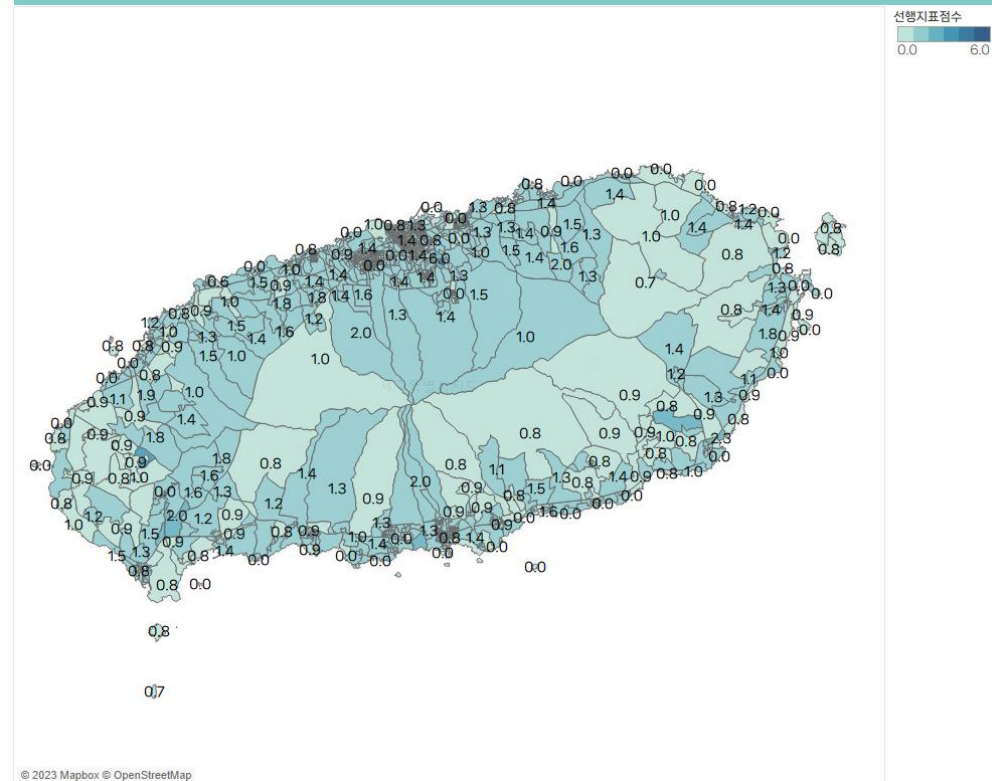


결론

집계구별 쇠퇴진단지표 종합점수



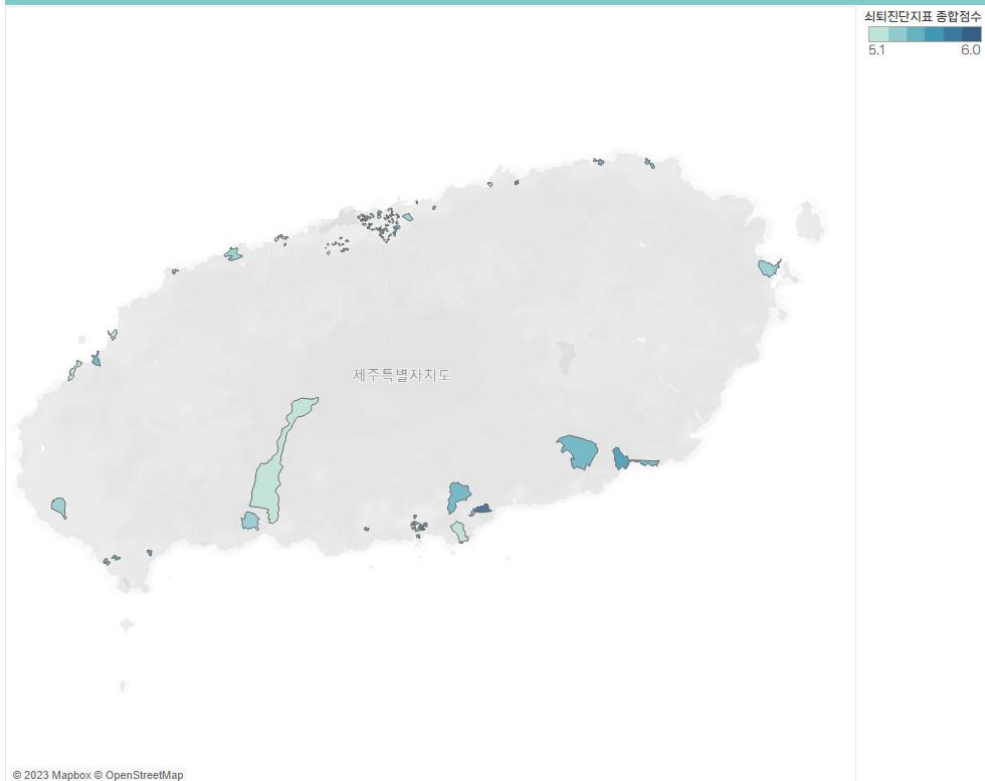
집계구별 선행지표점수



- 집계구별 쇠퇴진단지표 종합 점수와 선행 지표 점수를 검토하면서 현재 쇠퇴가 심한 지역과 미래의 쇠퇴가 예상되는 지역을 함께 고려하여 판단함
- 또한 주성분 분석 결과에서 37%, 35%의 분산 비율이 가지고 있는 1번 주성분의 주요성분과 2번 주성분의 주요 성분인 인구 수 데이터와 유동인구 데이터는 쇠퇴진단지표 점수에서 가중치를 고려 할 수 있음 (현재 각 지표마다 1점씩 부여함)

1 분석 결과

쇠퇴진단지표 종합점수 높은 순위 상위 100개 집계구



▼ 쇠퇴진단지표 종합점수가 높은 TOP10 지역

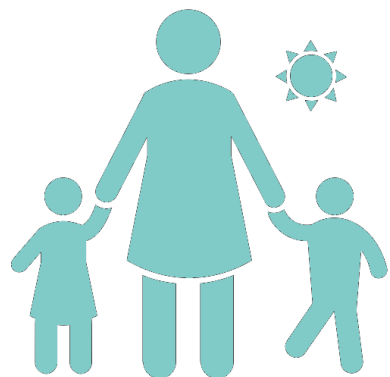
집계구 ID	39020530010007	39010540070602
	39020120010002	39010540010009
	39010600020005	39010520010032
	39010570010005	39010520010002
	39010540071901	39010140020024

선행지표점수 높은 순위 상위 100개 집계구



▼ 유소년/생산가능인구 수를 기반으로 한 선행지표점수가 높은 TOP10 지역

집계구 ID	39020590040006	39010630090002
	39020580011601	39010630020001
	39020570070003	39010600090003
	39020320050005	39010310020028
	39020130020101	39010140020014



- 주성분 분석을 통해 인구데이터가 도시의 쇠퇴에 큰 영향을 미친다는 것을 확인하였고 그 선행 지표인 유소년 인구와 생산 가능 인구를 중점으로 하여 도시재생사업의 추진에 도움이 되고자 함
- 따라서, 유소년 수를 늘리기 위한 정책을 도시재생사업에 반영하는 것은 미래의 지역별 쇠퇴도 개선을 위한 좋은 방향성이라고 생각함
- 예를 들어, 유소년을 위한 어린이집이나 유치원 지원정책 관련 사업을 유치하거나 1인 단위의 가족보다 다인 단위 가족의 이주를 유도하기 위한 주거 환경 개선이 필요
- 쇠퇴 점수가 높게 도달하기 전, 선행지표를 통해 쇠퇴를 예방하는 재생사업이 적절하게 이루어진다면 **인접 지역으로의 쇠퇴의 확산을 방지하고** 지속 가능한 도시재생이 가능할 것으로 예상됨

End of Documents.