

귀농·귀촌 생활 지수 개발 및
이를 기반한 **정착지 추천** 서비스

2018/8/30



목차

I 개발의 필요성 및 아이디어

II 데이터 수집, 전처리 및 데이터베이스 구축

III 귀농·귀촌 추천 시스템 환경 개발

VI 시사점, 한계 및 발전 방향



CHAPTER

I

개발의 필요성 및 아이디어



1. 개발의 필요성 및 아이디어

➤ 문제 인식

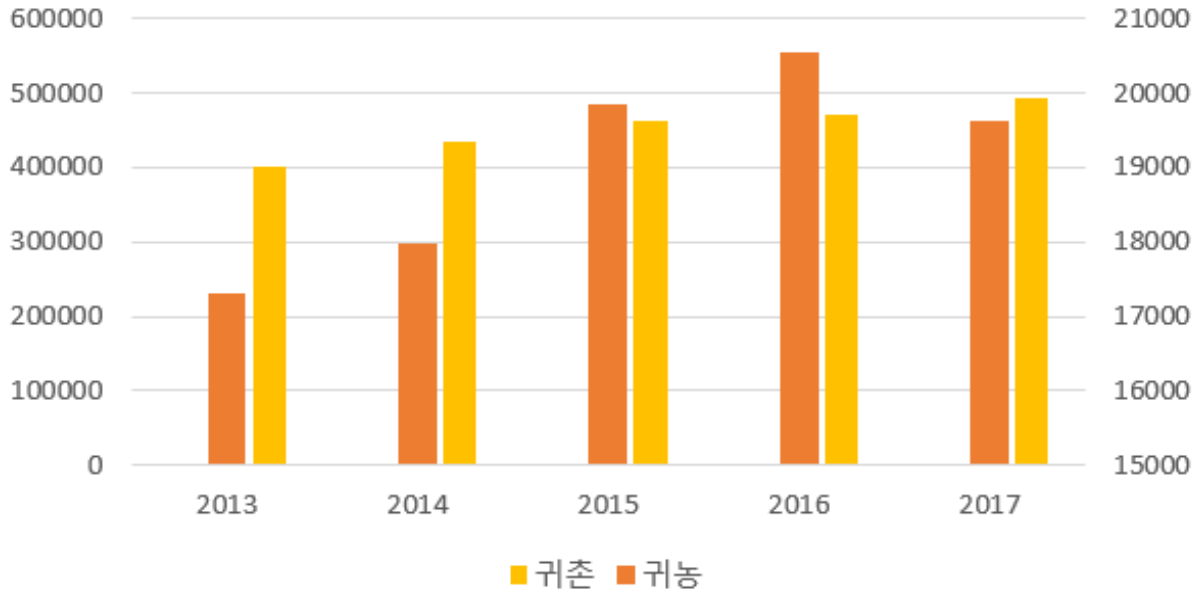
귀농·귀촌의 개념	귀농	농촌 이외의 지역에 살던 사람이 농업인이 되기 위해 농촌 지역으로 이주하는 행위
	귀촌	농촌 이외의 지역에 살던 사람이 거주를 목적으로 농촌 지역으로 이주하는 행위
귀농·귀촌 인구 지속적 증가		베이비붐 세대의 은퇴 및 탈도시 원하는 인구 증가
		지자체 및 정부의 지원 정책



1. 개발의 필요성 및 아이디어

➤ 문제 인식

최근 5년간 귀농·귀촌 인구수 변화 추이



2013년: 417,471 → 2017년 511,558 (22.5%)

귀농, 귀촌 인구 50만 돌파...40세 미만 젊은층 급증

28일 농림축산식품부가 발표한 자료에 따르면 2017년 귀농, 귀촌인 및 가구원은 총 51만6817명으로 집계됐으며 2013년 통계 작성을 시작한 이래로 처음으로 50만명을 넘은 것으로 조사됐다.

경남지역 상반기 귀농인구 122% 늘어

김해뉴스

24일 경남도에 따르면 올 상반기 경남으로 귀농을 신청한 가구는 651가구로 지난해(535가구)보다 122% 늘어났다. 이는 경남도가 올해 유치 목표로 세운 700가구의 93% 수준으로, 이대로라면 연말까지 목표치를 크게 초과할 전망이다.

[귀농귀촌 시대] 20대도, 30대도, 40대도... 농촌으로 고향으로

연합뉴스
YONHAP NEWS AGENCY

◇ 귀농·귀촌 가구 10년새 330배 급증

최근 농업과 농촌에 불고 있는 가장 눈에 띄는 변화는 귀농·귀촌 인구의 증가다.

27일 통계청에 따르면 2005년 1천여 가구에 불과했던 귀농·귀촌 가구는 2015년 약 33만 가구로 급증했고 이런 추세는 당분간 이어질 전망이다.



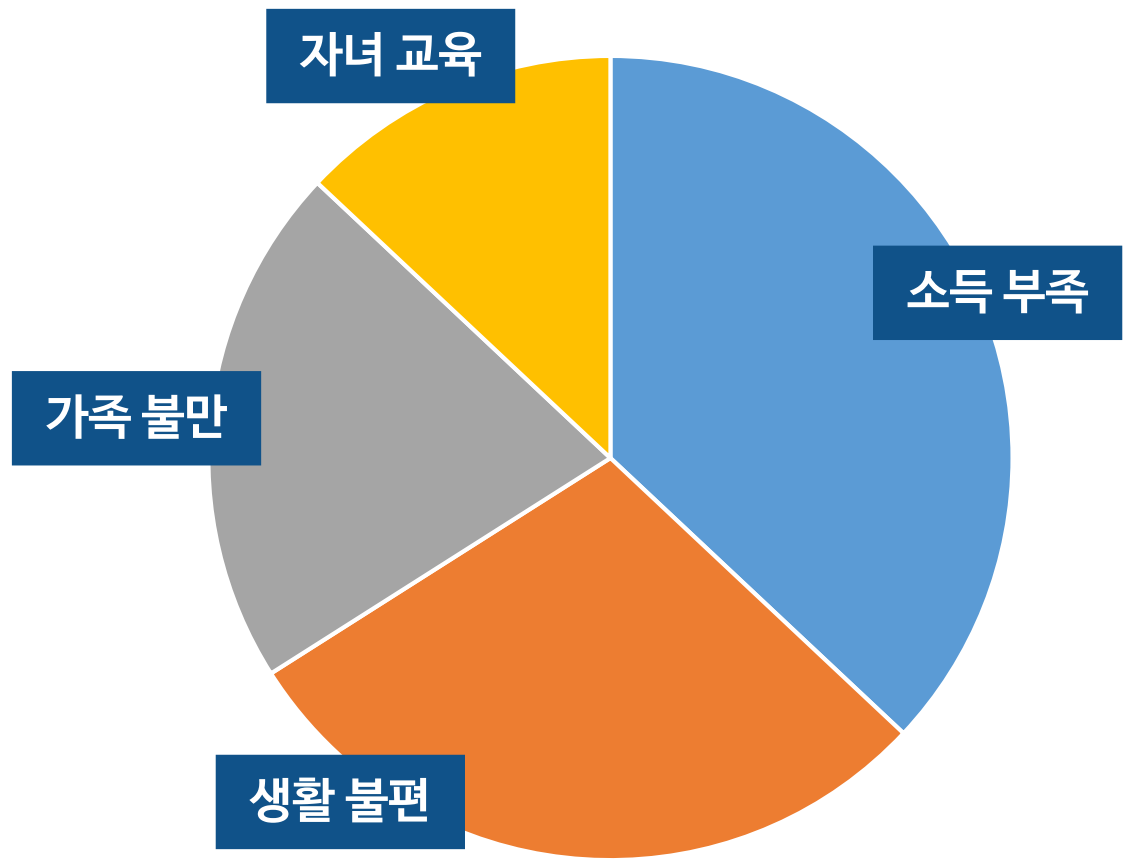
1. 개발의 필요성 및 아이디어

➤ 개발의 필요성

어설피게 덤벼들다간 낭패.. **귀농인 10명 중 1명 짐 싣는다**

연합뉴스 2017/02/04

지난해 농림축산식품부가 2012~2015년 귀농·귀촌한 1천가구를 대상으로 조사한 결과 농촌적응에 실패해 다시 도시로 되돌아오거나 계획 중인 경우는 각각 4%와 11.4%로 나타났다. 10명 중 1명은 다시 농촌을 등지고 역(逆) 귀농한다는 얘기가.



귀농·귀촌 희망자들의 목적에 맞는 정착지를 선택하는 것이 중요!



1. 개발의 필요성 및 아이디어

➤ 개발 아이디어

귀농·귀촌 추천 시스템 환경 개발

1

정착지 선정에 영향을 미치는 **주요 요소**를 파악

2

정착지로서의 매력도를 평가하는 **귀농·귀촌 생활 지수**를 정의

3

비슷한 생활 패턴을 지닌 **동일 생활권 지역** 추천



CHAPTER

II. 데이터 수집, 전처리 및 데이터베이스 구축



II. 데이터 수집, 전처리 및 데이터베이스 구축

➤ 데이터 전처리

시계열 데이터 -> 증가율 데이터로 변환

분석 복잡성 ↑
증가율이 더 의미 있는 종속변수
독립 변수 데이터 형식 통일 필요



연평균성장률 (CAGR, Compound Annual Growth Rate)

$$CAGR = \left(\frac{Final\ Value}{Starting\ Value} \right)^{\frac{1}{N}} - 1$$

독립변수 (지수계산용), 종속변수 정규화 작업

증가율: 데이터간 값 크기 고려 X



최소-최대 정규화 [0.01, 1]¹⁾

최소값 0.01, 최대값 1 사이의 값으로 데이터 정규화

1) 이현호, 'Python과 SQL을 활용한 실전 데이터 전처리', 2018.7, p.143 참고



II. 데이터 수집, 전처리 및 데이터베이스 구축

project(ORACLE)

PKFK	Name	Type
✓	sigungu_id	number(10)
	latitude	number(20,10)
	longitude	number(20,10)

PKFK	Name	Type
	sido	varchar(30)
	sigungu	varchar(30)
✓	sigungu_id	number(10)
	division	varchar(20)
	total_quotient	number(20,6)
	income_rate	number(20,6)
	avg_income	number(20,6)
	price_rate	number(20,6)
	avg_price	number(20,6)
	libraries	number(20,6)
	minor	number(20,6)
	school	number(20,6)
	theater	number(20,6)
	center_exist	number(1)
	metro	number(20,6)
	train_station_e	number(1)
	ktx_exist	number(1)
	work_distance	number(20,6)
	terminal	number(20,6)
	hospital	number(20,6)
	post_office	number(20,6)
	foreigner_rate	number(20,6)
	gender_rate	number(20,6)
	true_transfer	number(1)
	residential_are	number(20,6)
	official_price_r	number(20,6)
	official_price	number(20,6)
	old_house_rate	number(20,6)
	min_temperatu	number(20,6)
	max_temperatu	number(20,6)
	sum_precipitati	number(20,6)

PKFK	Name	Type
	sido	varchar(30)
	sigungu	varchar(30)
✓	sigungu_id	number(10)
	total_quotient	number(20,6)
	income_rate	number(20,6)
	avg_income	number(20,6)
	price_rate	number(20,6)
	avg_price	number(20,6)
	libraries	number(20,6)
	minor	number(20,6)
	school	number(20,6)
	theater	number(20,6)
	center_exist	number(1)
	metro	number(20,6)
	train_station_e	number(1)
	ktx_exist	number(1)
	work_distance	number(20,6)
	terminal	number(20,6)
	hospital	number(20,6)
	post_office	number(20,6)
	administrative	number(20,6)
	foreigner_rate	number(20,6)
	senior_rate	number(20,6)
	gender_rate	number(20,6)
	true_transfer	number(1)
	residential_are	number(20,6)
	official_price_r	number(20,6)
	official_price	number(20,6)
	old_house_rate	number(20,6)
	avg_temperatu	number(20,6)
	min_temperatu	number(20,6)
	max_temperatu	number(20,6)
	sum_precipitati	number(20,6)
	total_rate	number(20,6)
	under_fourty	number(20,6)
	fourty	number(20,6)
	fifty	number(20,6)
	over_sixty	number(20,6)
	total_rate_towr	number(20,6)
	under_fourty_t	number(20,6)
	fourty_town	number(20,6)
	fifty_town	number(20,6)
	over_sixty_tow	number(20,6)

PKFK	Name	Type
✓	sigungu_id	number(10)
	price_rate	number(20,6)
	income	number(20,6)
	official_price	number(20,6)
	total_populatio	number(20,6)

PKFK	Name	Type
✓	sigungu_id	number(10)
	price_rate	number(20,6)
	income	number(20,6)
	official_price	number(20,6)
	total_populatio	number(20,6)

PKFK	Name	Type
✓	sigungu_id	number(10)
	price_rate	number(20,6)
	income	number(20,6)
	official_price	number(20,6)
	total_populatio	number(20,6)

PKFK	Name	Type
✓	sigungu_id	number(10)
	price_rate	number(20,6)
	income	number(20,6)
	official_price	number(20,6)
	total_populatio	number(20,6)

PKFK	Name	Type
✓	sigungu_id	number(10)
	price_rate	number(20,6)
	income	number(20,6)
	official_price	number(20,6)
	total_populatio	number(20,6)

TOTAL : 총 취합한 데이터 테이블

DIAMETER : 위경도 데이터 테이블

CALC_TOTAL : 지수 계산용 취합 데이터 테이블

THIRTEEN ~ SEVENTEEN : 연도별 주요 4개 데이터 테이블



CHAPTER

Ⅲ

귀농·귀촌 추천 시스템 환경 개발



Ⅲ. 귀농·귀촌 추천 시스템 환경 개발

➤ 개발 과정





III-1. 정착지 선정 핵심변수 파악 (회귀분석)

회귀분석과정

단순회귀분석 -> 다중회귀분석 사용

	노년층비율	평균기온	행정기관1개당평균인구수
미성년자비율	-0.911381		
최저기온		0.91469	
우체국1개당평균인구수			0.987844

변수 선택 목적
단순회귀분석 결과
상관계수 값 너무 낮아
변수 선별 못함



OLS 사용
1차: 독립변수 29개
2차: 독립변수 26개
3차:
① 연령대: 39세 이하, 40대, 50대, 60세 이상
② 지역 구분: 남부/북부

다중공선성 진단

- 0.8이상이면 다중공선성 있다 판단
- 총 3개 변수 삭제

문제

- 70세 이상 p-value 0.1 이하 변수 부재
→ 60세 이상으로 수정
- 전반적으로 낮은 상관계수
→ 남부/북부 구분

잔차 분석
정규성 검증

〈단순회귀분석〉

〈다중회귀분석〉



III-1. 정착지 선정 핵심변수 파악 (회귀분석)

회귀분석과정

단순회귀분석 -> 다중회귀분석 사용

변수 선택 목적
단순회귀분석 결과
상관계수 값 너무 낮아
변수 선별 못함



OLS 사용
1차: 독립변수 29개
2차: 독립변수 26개
3차: ① 연령대: 39세 이하, 40대, 50대, 60세 이상 ② 지역 구분: 남부/북부

〈단순회귀분석〉

〈다중회귀분석〉

P> t	const 0.662
	연평균소득증가율 0.326
	5년평균소득 0.904
	연평균소비자물가변화율 0.685
	5년평균소비자물가 0.589
	공공도서관1개당인구수 0.560
	미성년자인구비율 0.613
	초중고 1개당 10~19세 평균인구수 0.461
	영화/극장 1개당 평균인구수 0.876
	귀농지원센터존재여부 0.862
	지하철역 1개당 평균인구수 0.754
	기차역존재여부 0.449
	KTX역존재여부 0.675
	인접도시통근거리 0.649
	터미널수 0.829
	병원1개당 평균인구수 0.363
	우체국1개당 평균인구수 0.787
	외국인비율 0.674
	성비 0.948
	순전입여부 0.546
	1인당 주거면적 0.858
	연평균공시지가 상승률 0.596
	공시지가 0.580
	노후주택비율 0.620
	최저기온 0.630
	최고기온 0.601
	합계강수량 0.894

문제

- 70세 이상 p-value 0.1 이하 변수 부재
→ 60세 이상으로 수정
- 전반적으로 낮은 상관계수
→ 남부/북부 구분

잔차 분석
정규성 검증

〈다중회귀분석〉



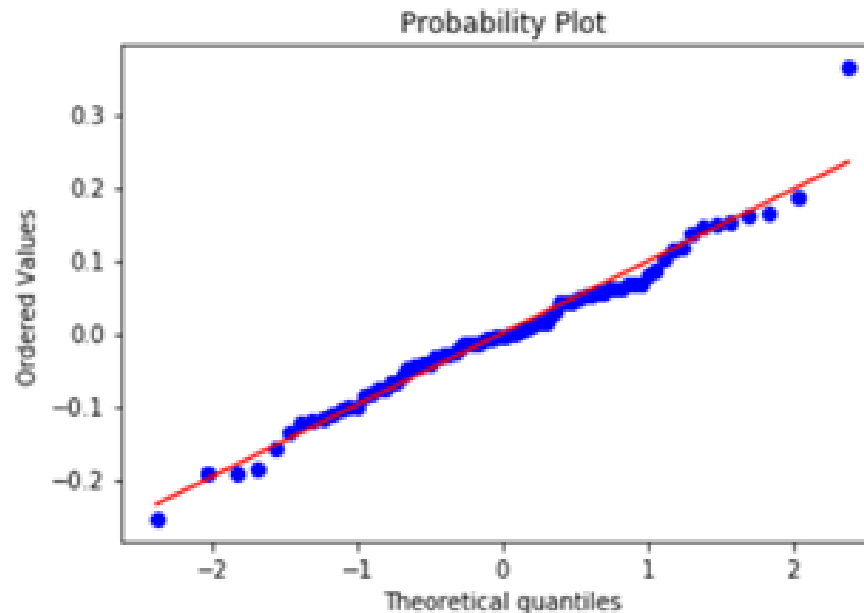
III-1. 정착지 선정 핵심변수 파악 (회귀분석)

회귀분석과정

단순회귀분석 -> 다중회귀분석 사용

#잔차 공산도

```
sp.stats.probplot(a_result.resid, plot=plt)
plt.show()
```



잔차 분석
정규성 검증

변수 선택 목적

단순회귀분석 결과
상관계수 값 너무 낮아
변수 선별 못함

OLS 사용

1차: 독립변수 29개

2차: 독립변수 26개

3차:

- ① **연령대**: 39세 이하, 40대, 50대, 60대 이상
- ② **지역 구분**: 남부/북부

〈단순회귀분석〉

〈다중회귀분석〉



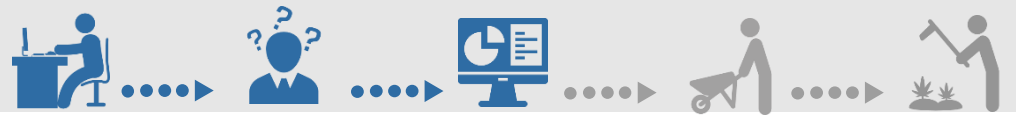
III-1. 정착지 선정 핵심변수 파악 (회귀분석)

귀촌

북부	39세 이하	1인당 주거면적(m2)	연평균 소득증가율	지하철역 당 평균인구 수	순전입 여부	미성년자 비율	기차역 존재여부
	40대	인접도시 통근거리	연평균 소득증가율	지하철역 당 평균인구 수	공시지가	미성년자 비율	KTX역 존재여부
	50대	공시지가	연평균 소득증가율	KTX역 존재여부	성비	미성년자 비율	연평균 공시지가 상승률
	60세 이상	공시지가	연평균 소득증가율	노후주택 비율	순전입 여부	성비	기차역 존재여부
남부	39세 이하	우체국 당 평균인구 수	연평균 소득증가율	연평균 소비자물가 변화율	순전입 여부	외국인 비율	연평균 공시지가 상승률
	40대	우체국 당 평균인구 수	연평균 소득증가율	연평균 소비자물가 변화율	순전입 여부	영화/극장 1개당 평균인구 수	연평균 공시지가 상승률
	50대	1인당 주거면적(m2)	연평균 소득증가율	연평균 소비자물가 변화율	순전입 여부	외국인 비율	연평균 공시지가 상승률
	60세 이상	1인당 주거면적(m2)	우체국 당 평균인구 수	인접도시 통근거리	순전입 여부	성비	영화/극장 1개당 평균인구 수

귀농

북부	39세 이하	공시지가	5년 평균소득	지하철역 당 평균인구 수	성비	연평균 소비자물가 변화율	연평균 공시지가 상승률
	40대	최저기온	5년 평균소득	귀농지원센터 존재여부	성비	연평균 소비자물가 변화율	기차역 존재여부
	50대	1인당 주거면적(m2)	5년 평균소득	귀농지원센터 존재여부	공시지가	연평균 소비자물가 변화율	우체국 당 평균인구 수
	60세 이상	공시지가	5년 평균소득	5년 평균 소비자물가	순전입 여부	연평균 공시지가 상승률	기차역 존재여부
남부	39세 이하	우체국 당 평균인구 수	공시지가	지하철 당 평균인구 수	순전입 여부	영화/극장 1개당 평균인구 수	연평균 공시지가 상승률
	40대	우체국 당 평균인구 수	최고기온	귀농지원센터 존재여부	순전입 여부	병원 1개 당 평균인구 수	외국인 비율
	50대	1인당 주거면적(m2)	연평균 소득증가율	인접도시 통근거리	최고기온	공시지가	지하철 역 당 평균인구 수
	60세 이상	1인당 주거면적(m2)	우체국 당 평균인구 수	5년 평균소득	공시지가	병원 1개 당 평균인구 수	연평균 소득증가율



Ⅲ. 귀농·귀촌 추천 시스템 환경 개발

➤ 개발 과정

Ⅲ 귀농·귀촌 추천 시스템 환경 개발

Ⅲ-1 정착지 선정에 영향을 미치는 **주요 요소**를 파악

Ⅲ-2 정착지로서의 매력도를 평가하는 **귀농·귀촌 생활 지수**를 정의

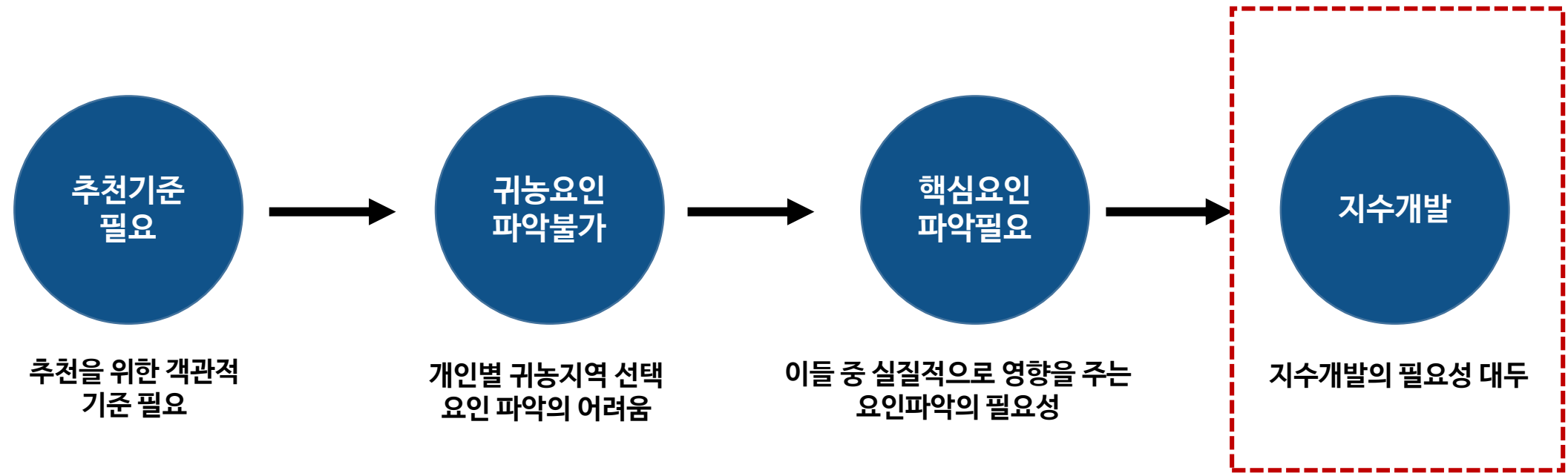
Ⅲ-3 비슷한 생활 패턴을 지닌 **동일 생활권 지역** 추천

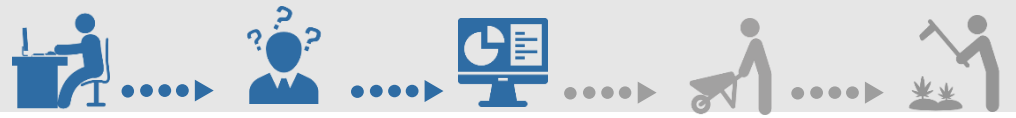


III-2. 귀농·귀촌 생활지수 개발

➤ 귀농, 귀촌 생활지수 도출

객관적인 추천 방법론의 필요성



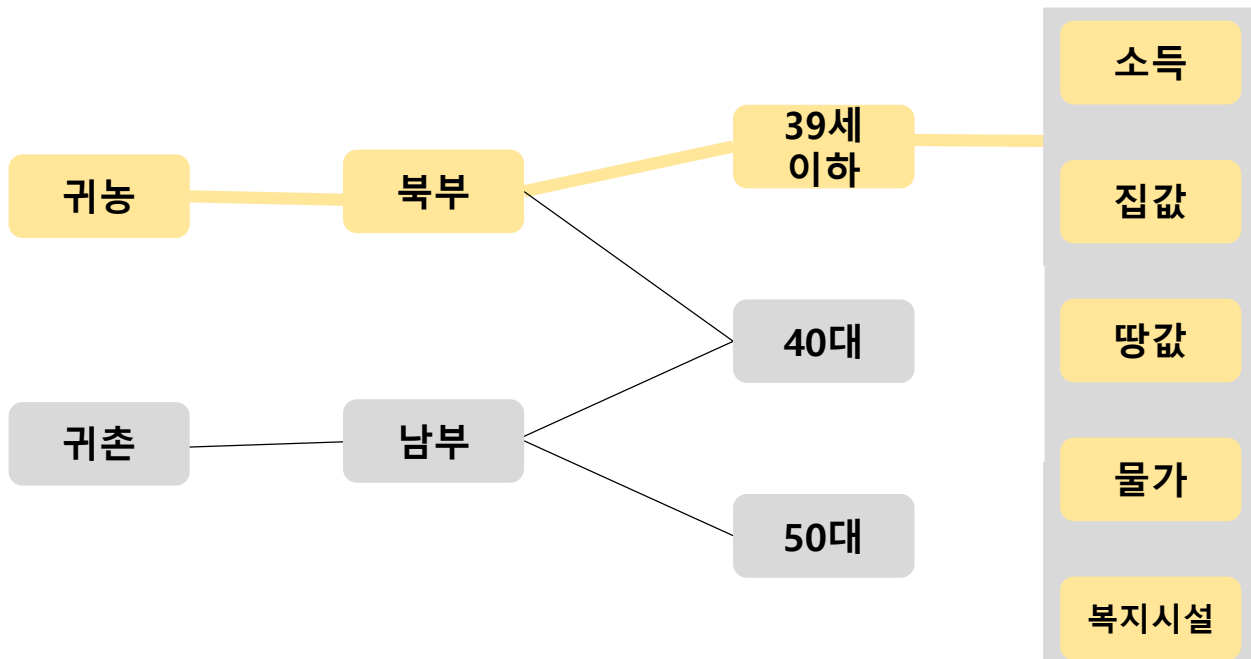


III-2. 귀농·귀촌 생활지수 개발

추천 지수 계산 원리



(맞춤형 핵심지수 X 지역별 실데이터)

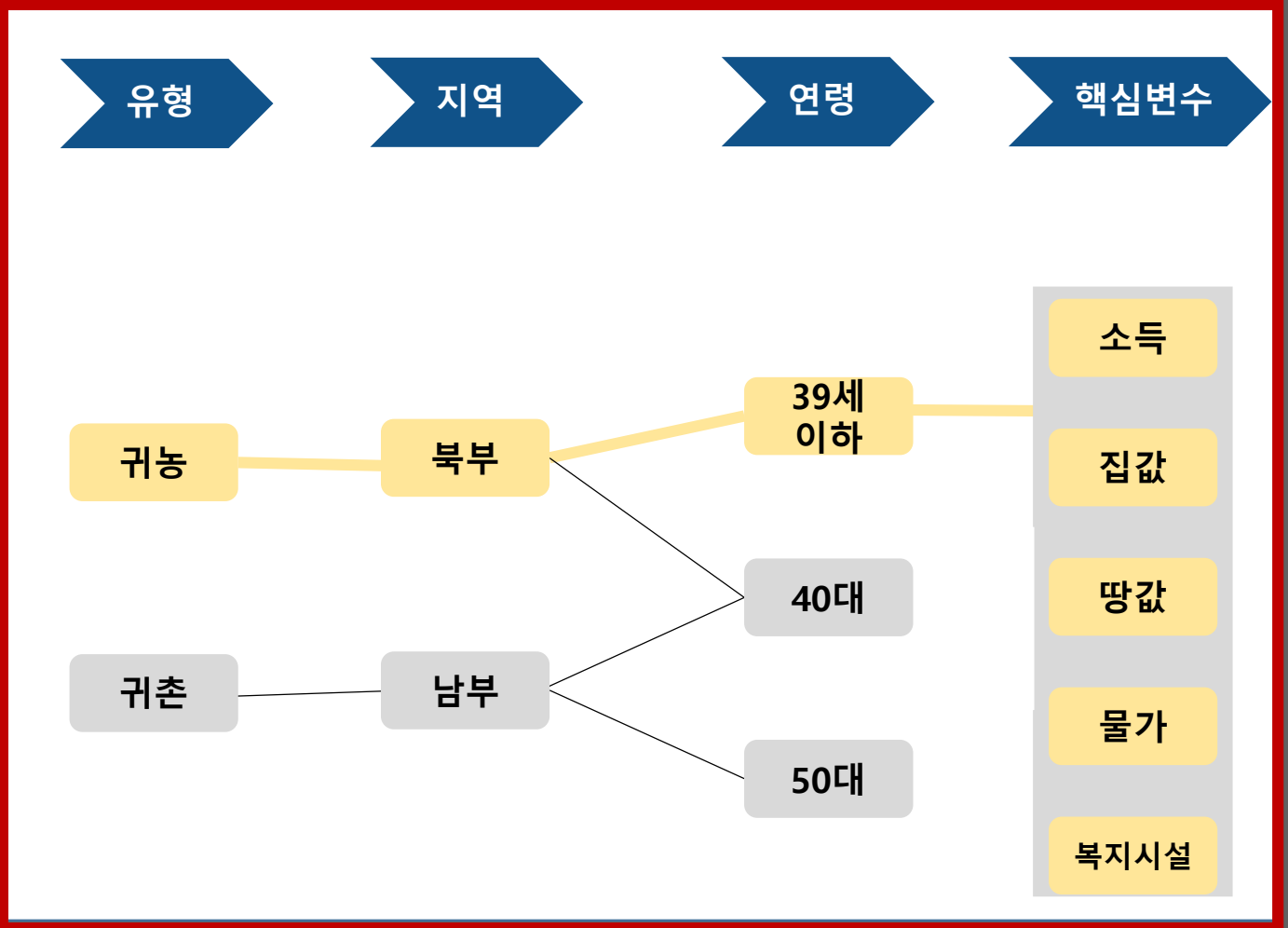


소득 상관계수	지역별 소득	순위	지역
✗	지역별 소득	1	곡성군
✗	지역별 집값	2	고흥군
✗	지역별 땅값, 물가..		
✗	지역별 땅값, 물가..		
✗	지역별 땅값, 물가..	229	고양시



III-2. 귀농·귀촌 생활지수 개발

추천 지수 계산 원리



실데이터 기준

맞춤형 지역추천

STEP 1

(맞춤형 핵심지수 X 지역별 실데이터)

카테고리별로 회귀분석을 통해
유형에 맞는 핵심변수 제공

소득 상관계수	지역별 소득 순위	지역
✖	지역별 집값	곡성군, 고흥군
✖	지역별 땅값, 물가..	
✖	지역별 땅값, 물가..	
✖	지역별 땅값, 물가..	229, 고양시



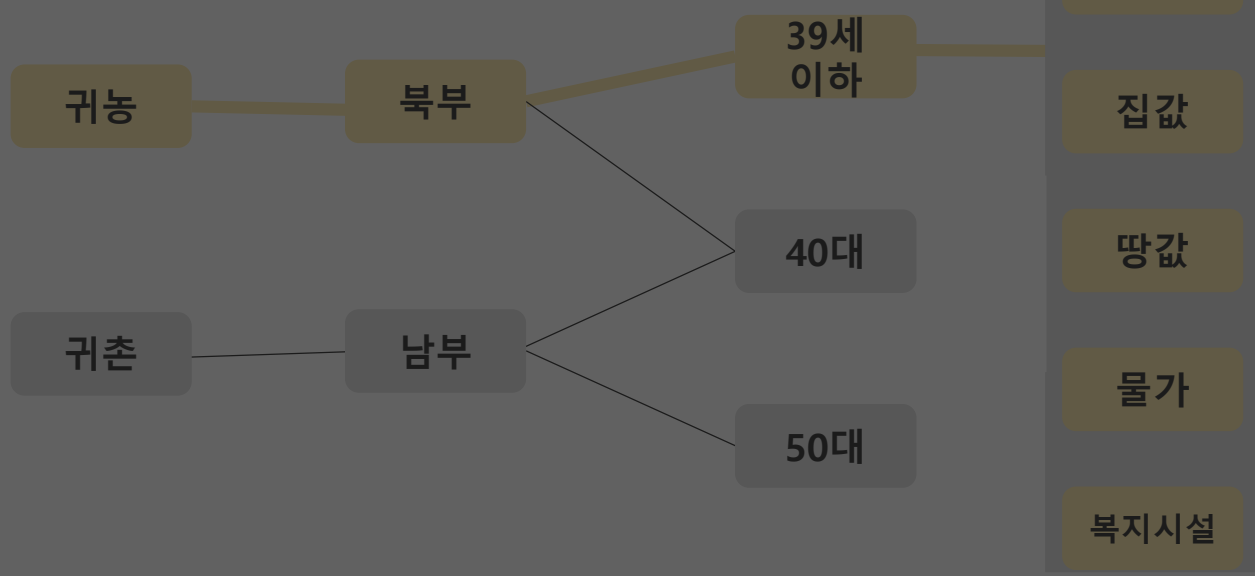
III-2. 귀농·귀촌 생활지수 개발

추천 지수 계산 원리



STEP 2

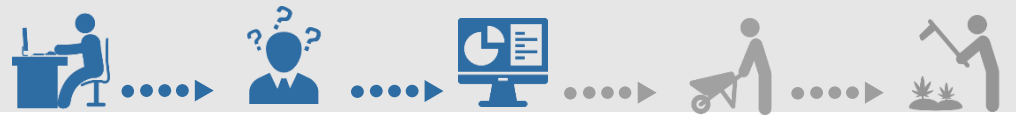
지수 = \sum (핵심변수(상관계수) X 실데이터 값)



실데이터 기준
맞춤형 지역추천

(맞춤형 핵심지수 X 지역별 실데이터)

상관계수	연산 결과	순위	지역
소득 상관계수	✗ 지역별 소득	1	곡성군
집값 상관계수	✗ 지역별 집값	2	고흥군
땅값 상관계수	✗ 지역별 땅값, 물가..		
물가 상관계수	✗ 지역별 땅값, 물가..		
복지시설 상관계수	✗ 지역별 땅값, 물가..	229	고양시

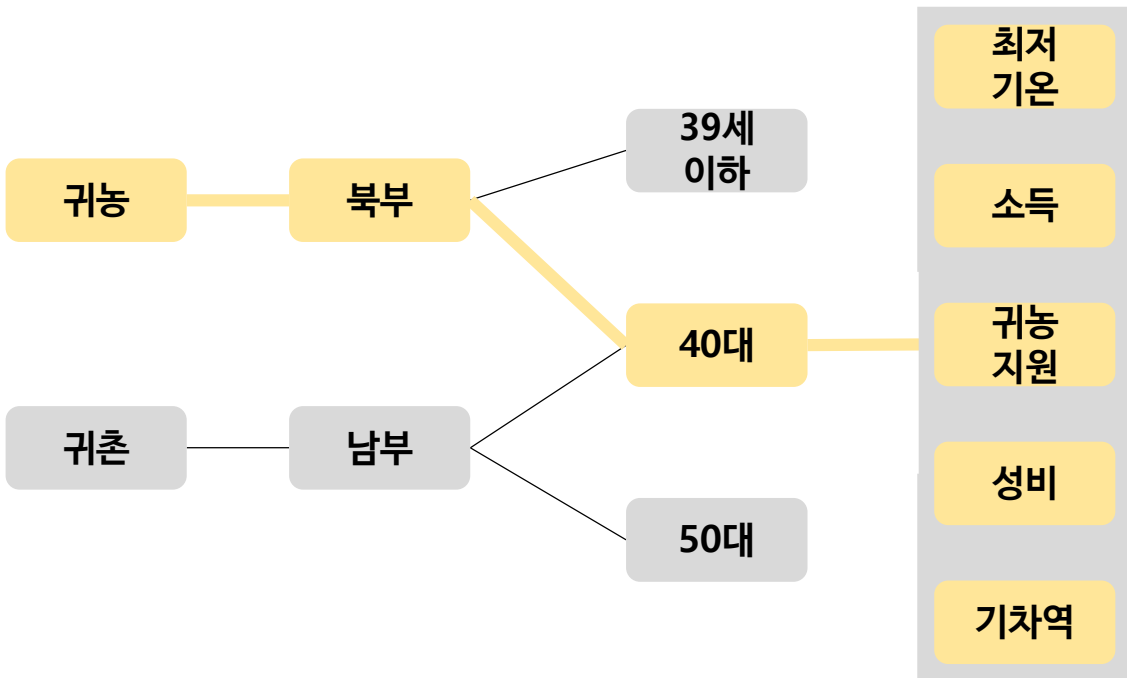


III-2. 귀농·귀촌 생활지수 개발

계산 예시



EX) 전라남도 나주시



핵심변수	귀농	귀촌	지수
최저기온	0.0167	0.512	0.0086
소득	-0.0005	0.239	0.0001
귀농지원	0.1511	1	0.1511
성비	0.8477	0.251	0.2128
기차역	-0.1739	1	-0.1739

※ 정규화[0.01,1]

$$\sum^* = 0.19$$

순위	지역
1	태안군
2	완주군
3	나주시
...	...
138	의령군



Ⅲ. 귀농·귀촌 추천 시스템 환경 개발

➤ 개발 과정





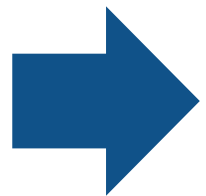
III-3. 동일 생활권 지역 추천 (클러스터링)

방법론

클러스터링 방법: K-means
클러스터 수: 10

시나리오 (예시)

선택 시군구: 부산광역시 기장군
탐색 특성: 1) 최저 기온
2) 외국인 비율
3) 노후주택 비율



	도시	최저기온	외국인 비율	노후주택비율
0	부산광역시 기장군	-7.5	0.012210	0.228174
1	대구광역시 달성군	-9.6	0.023023	0.296509
2	울산광역시 울주군	-7.4	0.032251	0.366981
3	전라남도 무안군	-6.9	0.019905	0.485284
4	경상북도 칠곡군	-10.2	0.031616	0.425238
5	경상남도 김해시	-8.4	0.034149	0.341528
6	경상남도 거제시	-6.8	0.028015	0.375510
7	경상남도 양산시	-8.5	0.017452	0.320529
8	제주특별자치도 제주시	-1.1	0.028138	0.433286
9	제주특별자치도 서귀포시	-1.6	0.042238	0.497827



CHAPTER

IV

시사점, 한계 및 발전 방향



IV. 시사점, 한계 및 발전 방향

➤ 기대 효과

귀농·귀촌 희망자

자신의 연령대/희망 지역/ 정착 타입에 맞는 정착지 추천 → **성공적인 정착 가능성 향상**

자신이 가고자 하는 지역과 비슷한 성향의 지역을 추천받을 수 있음 → **보다 많은 지역 검토 가능**

지방자치단체

지역이 가지고 있는 객관적인 귀농·귀촌지로서의 **매력도 파악 가능**

인기가 많은 정착지에 비해 부족한 점이 무엇인지를 인지 후 **보완 가능**

부동산·사업가

귀농·귀촌지로서 매력이 있는 지역에 경쟁 기업에 비해 앞서 **진출 가능**

정착지에 어떤 유형의 귀농·귀촌인들이 몰리는지 파악 가능



IV. 시사점, 한계 및 발전 방향

▶ 한계 및 발전 방향

데이터 튜플 부족

전국 시군구 중 농촌으로 분류되는 지역 138개 시군구를 대상으로 조사

시군구 단위가 아닌 읍면동 단위로 조사하여 보다 정확하고 방대한 데이터 확보의 필요성

정성적 데이터 고려 부족

독립 변수에서 정량적 데이터만을 사용, 이웃간의 관계 등 정성적 데이터는 고려되지 못함

정성적 데이터를 정량적 데이터로 해석할 수 있는 방법 마련/ 참고사항으로 제공

회귀분석·지수화 방법의 자의성

회귀분석 및 지수화 과정에서 학문적으로 확고한 방법론이 아닌 주관적인 판단이 개입됨

Lasso 회귀분석 등 보다 정교한 회귀방법론의 도입 필요성

교수님, 조교님, 담당자님, 학생 여러분 모두 감사합니다!

데이터 출처

공공 데이터 포털 - data.go.kr
기상청 - kma.go.kr
통계지리정보서비스 - sgis.kostat.go.kr
국가통계포털 - kosis.kr
네이버 지도 크롤링 - map.naver.com

참고 문헌

귀농·귀촌 컨설팅이 도시민의 농촌 이주 의
사결정에 미치는 영향에 관한 연구 - 권도
원, 2014
농촌 가서 뭐하고 살지? - 이진숙, 2013