# 데이터 수집계획 및 보고서

# [데이터 수집] 프로젝트 계획

프로젝트명	지역별 인구 이동 분석
능력단위분류	데이터 전처리 및 분석
수행 기간	2023.05.18 2023.05.26

### ■ 상세 정보

	설명 및 의견
데이터 수집 목적	및 연령별 인구 이동량을 파악하여 지역 간의 국토개발, 교통, 교육 및 주택 등의 각종 정책수립을 위한 자료로 활용할 수 있음
데이터 수집 계획	공공데이터포털에서 행정안전부_지역별 인구이동 현황 OpenAPI 수집 https://www.data.go.kr/tcs/dss/selectApiDataDetailView.do? publicDataPk=15108 093
데이터 수집 시스템 환경	Windows 11, Python 3.10.9, conda 23.1.0, pandas 2.0.1, matplotlib 3.7.1, seaborn 0.12.2
데이터 수집 후 저장 방법	Open API로 데이터를 수집한 후 data.json로 데이터 저장 데이터 정제 후 population_data.csv로 파일 저장
데이터 수집 후 정제 계획	불필요한 컬럼 삭제, 컬럼 타입 변경(NmprCnt가 포함된 컬럼 타입을 int로 변경), 컬럼 명 변경(영어를 한국말로), 컬럼 순서 변경, 남성과 여성을 연령대로 컬럼 수정(10~19세 는 10대와 같은 방식)

## [데이터 수집] 프로젝트 보고서

#### ■ 데이터 수집 작업

```
설명
              import requests
             amport requests
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import math
              import json
from pandas import DataFrame
              - open API 통해 2023년 1월부터 2023년 4월까지의 모든 데이터 추출
              headers = {
                 'accept': 'application/json',
              # 모든 페이지의 데이터 추출
             all response = []
              for y in range(202301, 202305):
                for i in range(math.ceil(289/100)):
                    params = {
                         'serviceKey': 'l4LZrZuxATppYOmrEIW59Ezozth1Yop8M2a7lsi2lpd/v+/0t6WglLJXQ2Am/ladh9L0y/6Ach7Xy4km6WF8LA==',
                         'mvinAdmmCd': '10000000000',
                         'mvtAdmmCd': '1000000000',
                         'srchFrYm': y,
                         'srchToYm': y,
                         'lv': '1',
                         'type': 'JSON',
                         'numOfRows': '100',
                         'pageNo': i+1
코딩
                    response = requests.get('https://apis.data.go.kr/1741000/ppltnDataStus/selectPpltnDataStus', params=params, headers=headers)
                    result = response.ison()
                    all response.append(result)
               · data.json 파일로 저장
              with open('data.json', 'w') as f:
json.dump(all_response, f)
               data.json 파일 불러오기
              with open('data.json', 'r') as file:

    raw_data = json.load(file)

    raw_data
```

구분					설명										
	<pre>item_list = [] for in raw_data;</pre>														
	df = DataFrame(Item_list) df														
	male4	AgeNmprCnt ma	le76AgeNmprCnt	male31AgeNmprCnt	feml55AgeNmprCnt	mvtCtpvNm	feml47AgeNmprCnt	male84AgeNmprCnt	male68AgeNmprCnt						
	0	143	71	985	242	서울특별	359	19	130						
	1	1	2	29	2	부산광역 시	4	0	1						
	2	2	0	12	1	대구광역	3	1	0						
	3	3	1	46	9	인천광역	12	0	7						
	4	0	0	12		광주광역	6	0	1						
						A									
	1151	1	9	2	1	전라복도	2	0	0						
	1152	1	1	2	o	전라남도	1	0	0						
	1153	0	0	3	1	검상복도	2	0	1						
	1154	0	0	1	2	경상남도 제주특별	0	0	0						
	1155	12	0	48	28	"자치도	19	1	14						
기타 설명 및 의견	있어 연령 데이터를 INVALIC 에서 파 알게 되었	병별로도 ( · 수집하는 )_REQUE: 라미터에 ! 었습니다. 하여 모든	기동량을 일 과정에서 ST_PARAN 대한 자세현 또한 모든	수 있습니다 계속 데이터 METER_ERR 한 설명이 부 페이지를 추	[구 이동을 약 나. 가 추출되지 OR 에러가 ' !족하여 파라 출하는 과정 정과 4개월 대	않고 발생하여 미터를 에서 계	원인을 알( 잘못 설정해 속 데이터 1	아내는데 힘들 오류가 발생 00개만 추출	들었으나 자료 행했다는 것을 ·되어서 for5						

#### ■ 데이터 수집 결과

구분	설명
	- 불필요한 컬럼 삭제
	# 컬럼 삭제
	df.drop(columns=['mvinAdmmCd', 'mvtAdmmCd', 'mvinDongNm', 'mvtDongNm', 'mvinSggNm', 'mvtSggNm'],
	inplace=True) - 컬럼 타입 변경
데이터 수집 후	# 컬럼 타입 변경  type_change = df.columns[df.columns.str.contains('NmprCnt')]  df[type_change] = df[type_change].astype(int)  df['statsYm'] = df['statsYm'].astype(int)  df.dtypes
정제 결과	male4AgeNmprCnt int32 male76AgeNmprCnt int32 male31AgeNmprCnt int32 feml55AgeNmprCnt int32 mvtCtpvNm object
	male60AgeNmprCnt int32 male103AgeNmprCnt int32 feml83AgeNmprCnt int32 feml70AgeNmprCnt int32 feml96AgeNmprCnt int32 Length: 234, dtype: object - 컬럼명 변경

									î	명										
# 2																				
df.	rena	me(	col	umns																
							inCt	· Charles				100	177							
							tCtp							٠,						
							tNmp				F-17-17-17	Land.	-							
							LeNm													
			¥80000	•			n LNm	prC	nt'	: "	여성	형인	구수	- ' } ,	,					
_	人	서 변	inp	Lace	e=Tr	ue	)													
# <i>캠핑</i>	_	. –	6																	
male_ag female_	e_colum	ns = [																		
									age 1	n re	nge(o,	111)]								
new_col df = df									인구수	.,	부성인-	구수']	• new_	column,	order]					
df[:5]																				
		*** DOOR																		
		전 전 입 출	총인	방	성 (	여성														
- d		시 시도 도	구수		구 · ( 수	인구 수	male0A	geNmpr	Cnt	male	1AgeN	nprCnt	mal	e2AgeN	mprCnt	mate	e3AgeN	mprCnt		
	- 0	명 명																		_
		서 서 울 울																		
0 20	7	투 투 별 별	54618	B 263	47 28	271			157			150	)		139			142		
		시시																		
		서 부 울 산																		
	2301	투 광	1675	5 8	23	852			2				X .		0	60		19		
1 20		별역				7-35			6			6	2						1000	
		별 역시 시					미N	lmnn		Τ:	하 컨								1 1070	
- 남	넑, 여	변 역 시 시  성 인					및 N	lmpr		포	함 컬								1 1000	
- H	성, 여 <sup>명함대</sup> :	별 역 시 시  성 인					및 N	lmpr		포	함 컬								1 1041	
- 남성 (설립, 역립	성, 여 5분대로 호 = [] = []	별 역 시 시  성 인					및 N	mpr		포	함 컬								1 1000	
# W.W. Orders female_columns female_	5, 0 500		면령대 break(stark)	로 킬			및 N	Impr		포	함 컬							4	1 1020	
# 150, 000 male, column for i in range start = 1 male, column for i in range start = 1 male, column famile, col	5, 0 = 0 = 0 = 0 = 0 = 0 = 0 = 0 =	별 역 시 시 인 # 4 # # # # # # # # # # # # # # # # #	변명대 breat(start) -forest(start	로 컬	id 4	수정	Cor , formati	end)].sum(e	Cnt	포	함 컬							4 22	1 3020	
# 5/3, 0/3/ rais, column for i in non start = end = i male, column famile, column famile, column famile, column famile, column famile, column	58, 0 58, 5 = [] 10, 5, 15 1 10, 15, 15 11 10, 15, 15 10, 15	별 역사 시 시성 인 화 수경 대 임역 여성 하나 (10년 ) 이 이 이 이 이 이 이 이 이 이 이 이 이 이 이 이 이 이	변명대 breat(start) -forest(start	로 컬	id 4	수정		end)].sum(e	Cnt	포	함 컬								1 1000	
# 158, 000 male, column for 1 in race start = and = i male, column for the race start = and = column for the race, column for the race column for	STORE S  STORE S  IN = []  IN = []  IN = []  IN TARRE = []  IN TAR	변 역 시 시 OL 항 수경 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	변경대 breat(start) .foreat(start la[]kpeller feet(]kpeller	로 耆	iden) 'al (pan) 'al	수정 ellipeter erlipete	Cor , format ( print , forma	end)].sum(e	Cnt	王	함 컬								1 1000	
a gra, organisa, columno female, columno femal	Signa S = [] ta = [	별 역 시 시 이 이 명	변경대 breat(start) .foreat(start la[]kpeller feet(]kpeller	로 耆	iden) 'al (pan) 'al	수정 ellipeter erlipete	Cor , format ( print , forma	end)].sum(e	Cnt	포	함 컬								1 1200	
# Stg. Office selections of the selection of the selectio	Signa s  = []  = [	변 역 시 OL # ## ## ## ## ## ## ## ## ## ## ## ##	변경대 breat(start) .foreat(start laf[]kpeller foot()kpeller(st 1000kpeller(st 1000kpeller(st	로 耆	iden) 'al (pan) 'al	수정 ellipeter erlipete	Cor , format ( print , forma	end)].sum(e	Cnt	平	함 컬								1 1200	
a Way, Ordinate femile, column for 1 in read of male, column for 1 in read of male, column femile, column femile, column femile, column femile, column a MONTER of femile of femile of femile of femile of femile of femile of	Signification of the control of the	변 역 시 OL # ## ## ## ## ## ## ## ## ## ## ## ##	변경대 breat(start) .foreat(start laf[]kpeller foot()kpeller(st 1000kpeller(st 1000kpeller(st	로 耆	iden) 'al (pan) 'al	수정 ellipeter erlipete	Cor , format ( print , forma	end)].sum(e	Cnt	平	함 컬								1 1200	
# 15 % And make, columns female, columns femal	STATE TO THE PROPERTY OF THE P	별 역시 OC 항수항  [작 남병 : f   [약 대명 : f   [약 ] ]   [약 대명 : f   [약 ] ]   [약 대명 : f   [약 ]   [약	면령대  format(start)  format(start)  format(start)  format(start)  format(start)	로 콘	estert): "all c(start): "al c(start): " eleptor ], Apiero (st	A. A. C.	Cor .format( porCor .forma (3)	end)].sum(e	Cnt	王	함 컬							4		
# STE, Order to start	Signa	변 역 시 O L  10	HESCH  tomat(start)  format(start)  format(start)  format(start)  format(start)  format(start)  format(start)  format(start)	로 콘	estert): "all c(start): "al c(start): " eleptor ], Apiero (st	A. A. C.	Cor .format( porCor .forma (3)	end)].sum(e	Cnt	포	함 컬									
# Mrg. Originals, columns female, columns # Mrg. Female, columns # Mrg. Female, columns female	Signal B = 0	행 역 시 O L  10	HESCH  tomat(start)  format(start)  format(start)  format(start)  format(start)  format(start)  format(start)  format(start)	로 콘	estert): "all c(start): "al c(start): " eleptor ], Apiero (st	A. A. C.	Cor .format( porCor .forma (3)	end)].sum(e	Cnt	포	함 컬									
* 192, de sala, colores ferals, colores ferals, colores colores (colores ferals, colores ferals, colores (colores ferals, colores (colores ferals, colores ferals, colores ferals, colores ferals, colores ferals, colores (colores ferals, colores ferals, colores ferals, colores ferals, colores ferals, colores (colores ferals, colores ferals, color	STORE S = [] = [] = [] = [] = [] = [] = [] =	間 M A O L	면명대 format(start) format(start) format(start) for (jugan format(s) for in the start)	로 콘  t) for increase in range(i) in range(i) in range(i)	estert): "all c(start): "al c(start): " eleptor ], Apiero (st	A. A. C.	Cor .format( porCor .forma (3)	end)].sum(e	Cnt	포	함 컬								1 1000	
্ প্রস্তুত্ব প্রস্তুত্ব করি, বাংলাকের করিব	Signal Si	행 시 O L  100 여명 시 O L  100 여명 이명 N O L  100 여명 이명 N O L  100 여명 N O L  100 어명 N	Ted CH  terracitate()  formaticate  formatic	로 콘 t) for iformat format in range(i) in range(i) in range(ii)	edd 2 ((tart): "d (tart): " (tart): " (tart):	A. A. C.	Cor .format( porCor .forma (3)	end)].sum(e	Cnt	포	함 컬							•		
* 1 Sign and set of the set of th	Signal Si	행 시 O L  100 여명 시 O L  100 여명 이명 N O L  100 여명 이명 N O L  100 여명 N O L  100 어명 N	Ted CH  terracitate()  formaticate  formatic	로 콘 t) for iformat format in range(i) in range(i) in range(ii)	edd 2 ((tart): "d (tart): " (tart): " (tart):	A. A. C.	Cor .format( porCor .forma (3)	end)].sum(e	Cnt	포	함 컬				3			,		
# 30%, 82%, 82%, 82%, 82%, 82%, 82%, 82%, 82	STOCK S = 0 = 0 = 0 = 0 = 0 = 0 = 0 = 0 = 0 =	행 시 O L  마 보험 :   마	Ted CH  terracitate()  formaticate  formatic	로 본 기 (formation formation formatio	[ [ [ [ [ [ [ [ [ [ [ [ [ [ [ [ [ [ [	AZA CO	Comp. (Spread)	end)].un(e	Cnt		16य व	램	· 나게	all of such	ડાંઘુ લ	अप थे	अब व व	sa pas	અવ ળ	
্য গ্ৰন্থ , ব্যক্তি , ব্যক্তি , ব্যক্তি ।  ক্ষেত্ৰ নিৰ্দেশ্য কৰি ।  ক	Signal Si	盟 AO L  G A A B  G A A B  G A B A B  G A B A B  G A B A B B A B B B B B B B B B B B B B	Terracissari)  formacissari for	文 (c) formation for a formation formation for a formation formation for a formation formation for a for	ister): will interest to the control of the control	A. A. C.	Corr. former()  formal former()  formal form	end)].sum(e	Cnt			럼 4	<b>낚제</b>	44 <b>14 o</b> 8' 3818 S			70대 대 성 856			
# 500, mg miss, others for it in most continued to the most contin	50, 00  SSURE 5 = 11  SSURE 5 = 12  SSURE 5	발 에 시 O L	中では、 ・	로 론 기 (formation formation formatio	(start)   mail (st	는 I MARIA M	Sor forests from the state of t	and), surface)], surface)], surface)], surface)], surface)], surface)], surface)	Cnt	117	164 d	24택 역 경	각제 <sup>30대 영</sup> 생	성	5세 역 경	श्चा थ स	성	क्रम् व स	અવ જ	190
# \$20, 025 and a second production of the control o	20	범 시시 C 로 하는 보급 (C C C C C C C C C C C C C C C C C C C	면령다  termac(start)  formac(start)	로 콜 로 로 로 로 로 로 로 로 로 로 로 로 로 로 로 로 로 로	del   -	aufligetones auflei pelo esti figure esti	Sor forests from the state of t	204 2 g g g g g g g g g g g g g g g g g g	Cnt	***	1806 g 1806 gl	18년 년 1837	38대 역 생 8455	성 3818	अपा जु व	ागा श्री व	896 896	88E 9 8 42	944 of 69	11

- 5 -

구분	설명
	- 정제한 데이터 population_data.csv로 저장  df.to_csv( population_data.csv', encoding='aff-8', index-False)  df = pd.read csv('population data.csv')
	선([:2]) 전입 전호 총인 남성인 여성인 이대 10대 20대 30대 10대 20대 30대 40대 50대 60대 70대 80대 대 대 대 대 대 대 대 대 대 대 대 대 대 대 대 대 대
	서울 서울 0 202301 특별 특별 54618 26347 28271 1717 1968 6343 6666 1898 8373 6465 3018 2695 1911 896 442 99 시 시
	서울 부산 1 202301 특별 광역 1675 823 852 18 45 430 152 66 458 126 71 53 30 16 11 0 ( 시 시
	2 rows x 28 columns
분석 결과 (의견)	불필요한 컬럼 삭제, 컬럼 타입 변경, 컬럼명 변경, 컬럼 순서 변경, 남성과 여성을 연령대로 컬럼 수정을 통해 정제된 데이터는 1156행 28열로 이루어져 있습니다.

- 6 -