

데이터 시각화를 통해 알아보는 부산 사립대학들의 미래 전망 예측

Team 4024

(40이서 0원히 2날을 𣦵때까지 기억하자)

팀장 권수현
팀원 김채영
박형준
이혁재

Contents

프로젝트 개요

- 프로젝트 목적
- 주제 선정 배경
- Data Set
- 개발 환경
- 역할분담

분석 설계

- 가설 설정
- 탐색적 데이터 분석
- 데이터 전처리
 - 데이터 정규화
 - 결측치 처리
 - 이상치 처리

데이터 분석

- 데이터 분석
- 데이터 시각화

결론 및 논의

- 결론
- 한계 및 발전사항
- Q & A

Google Colab을
활용한 Python 데이터
분석 역량 강화

Python 활용

Matplotlib &
Seaborn을 활용한
데이터 시각화의
활용 역량 강화

데이터 시각화

데이터를 기반으로 구체
적인 문제를 정의하고
해결하는 경험을 통해
문제해결능력 강화

문제 해결 능력

부산광역시에 위치한
4년제 사립대학교들은
학령인구 감소와
재정난으로 인해 생존에
심각한 위기를 겪고 있습니다.

본 프로젝트는
이러한 위기의 근본 원인을
데이터 분석을 통해
체계적으로 파악하고,
데이터 시각화를 활용하여
미래를 예측함으로써
부산 지역 4년제 사립대학교들의
지속 가능성을 확보하기 위한
현실적이고 실행 가능한 해결책을
제시하는 것을 목표로 합니다

YTN PICK · 2024.02.18. · 네이버뉴스

지방대 폐교 도미노 시작..."2040년 절반 이상 사라질 위기"

이 추세대로라면 2040년에 지방대 절반 이상이 사라질 수 있다는 분석까지 나왔습니다. 보도에 송세혁 기자입니다. [기자] 1995년 문을 연 강원관 광대학교. 학생은 한 명도 보이지 않고 건물은 텅 비었습니다. 신입생 급...



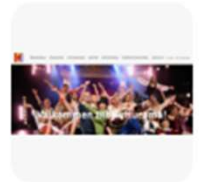
지방대 폐교 도미노 시작..."2040년 절반 이상 사라질 위기" YTN · 2024.02.18. · 네이버뉴스

[자막뉴스] 심각한 지방대 현실...우범지대가 된 캠퍼스 YTN · 2024.02.18. · 네이버뉴스

뉴스핍 · 2024.06.22.

[최연혁 교수의 '이제는 정치혁신' 대한민국의 급소, 지방경쟁력

2040년까지 전남에 있는 사립대학들의 생존율은 19퍼센트, 경남은 21퍼센트, 부산도 30퍼센트에 불과할 것이라고 한다. 현 정부들여 '교육발전 특구'와 '글로벌대학30' 등 지방 교육과 지방대 활성화를 위한 정책을 도...



매일경제 · A27면 1단 · 2023.10.17. · 네이버뉴스

지방대 '벚꽃엔딩' 증명 2040년 절반 사라질 듯

저출산으로 인한 학령인구 감소, 청년층 수도권 집중 심화 등으로 인해 '벚꽃 피는 순서대로 소멸될 수 있다'는 지방대학의 '벚꽃 엔딩' 시나리오가... 2022년 출생아 수 25만명, 대학 입학 정원 47만명이라는 수준이 계속 유지된다는 전제하에 2040년 초에는 50% 이상의 ...

데이터 명	형태	내용	출처	출처
2014 – 2024 년 고등 학교별X학과별 입학정원 지원 입학 학생 외국인학생 졸업 교원	CSV	각 년도별 학교/학과별 상반기 신입/재적/졸업/유학/ 교원 수	교육통계 서비스	https://kess.kedi.re.kr/contents/dataset?itemCode=04&menuId=m_02_04_03_02&tabId=m2
2020 – 2023 년 해외한류실태조사 보고서 요약본	PDF	각 년도별 국가별 한류콘텐츠 소비자의 한류 인식, 한류 소비 행태 동향 분석	문화체육관광부 · 한국국제문화교류 진흥원	https://kofice.or.kr/b20industry/b20_industry_00_vie.asp?seq=1049&page=1&find=&search=
2015~2023 년 연령_및_성별_인구 _읍면동	CSV	각 년도별 전국 인구수	국가통계 포털	https://kosis.kr/statHtml/statHtml.do?sso=ok&retur nurl=https%3A%2F%2Fkosis.kr%3A443%2FstatHt ml%2FstatHtml.do%3Flist_id%3D350_35001_B% 26obj_var_id%3D%26seqNo%3D%26tblId%3D% 35001_A009%26vw_cd%3D%26MT_ZTITLE%26orgl d%3D%26path%3D%252FstatisticsList%252F statisticsListIndex.do%26conn_path%3D%26MT_ZTIT LE%26itm_id%3D%26lang_mode%3Dko%26scrld %3D%26

운영체제

Windows 10
Mac Os 13.6.1

개발환경

Google Colab

언어

Python 3.13.1

데이터 분석

Numpy(1.26.4)
Pandas(2.2.2)
Scipy(1.13.1)

시각화

Matplotlib(3.8.0)
Seaborn(0.13.2)

Part 1

역할분담 및 프로세스

권수현

프로젝트 총괄 리딩 및 아이디어 수립
시각화 도구와 분석 기법 추천
문제 해결 컨설팅 및
팀 내 커뮤니케이션 활성화
발표 자료 제작 및 팀 조율

김채영

서브 데이터 전처리
데이터 시각화
분석 근거 자료 수집
해결 방안 모색

박형준

데이터 조사 및 효율적인 코드 작성
데이터 시각화
동적 그래프 생성
데이터 전처리 지원

이혁재

데이터 및 근거 자료 조사
가설설정
데이터 시각화

분석설계: 가설 설정

전국 학령인구를
감소로 인해
부산에 위치한
4년제 사립대학교에
대한 선호도가
감소하고 있다.

가설 1

부산의 4년제 사립대학교들은
학령인구 감소로 국내 학생 확보가
어려운 상황입니다.
하지만, 한류 콘텐츠를 활용한 국제적
홍보와 유망 학과 개설을 통해
유학생을 유치한다면,
충원율 문제를 해결하며 지속 가능한
발전을 이룰 수 있을 것입니다

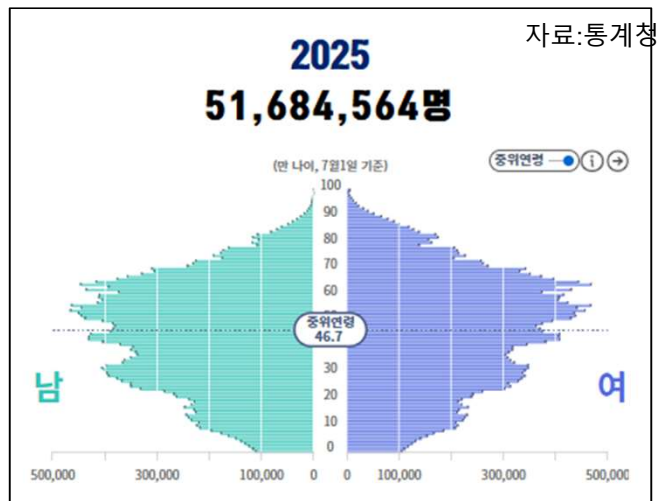
가설 2

Part 2

가설설계

:출산율 감소로 인해 학령인구 감소로 부산 사립 대학 존폐에 영향을 미칠 것이다.

[2025 인구 구성도]



출산율 감소에 따라 대학교에
진학하고자 하는 연령층이 감소되었음

[부산 사립대의 위기 관련 기사]

연합뉴스 · www.yna.co.kr > 연합뉴스 > 최신기사

부산 사립대총장 "위기극복 위해 지원책 필요·경쟁력 키워야" | ...

2021.06.15. (부산=연합뉴스) 김선호 기자 = 입학생 감소로 위기를 겪는 **부산지역 사립대** 총장들이 부산시에 산학협력 등 제도적 지원과 함께 스스로 경쟁력을 ...

천지일보 · www.newscj.com > news

'절체절명 폐교 위기' 부산지역 대학 생존을 겨우 30% 전망

2021.12.07. 사립대학, 엄연한 지역 자산응급주사 아닌 장기대책 절실대
학 간 통폐합, 경쟁력 강화상인 "폐교로 매출 반 토막"제각각 청년정책 보
완 필요시 "고용 안정화 모색할 것"[천지일보 부산=윤선영 기자] 수능이...

서울과 거리가 있고 제 2의 수도인
부산의 대학은 위기에 놓여 있음

Part 2

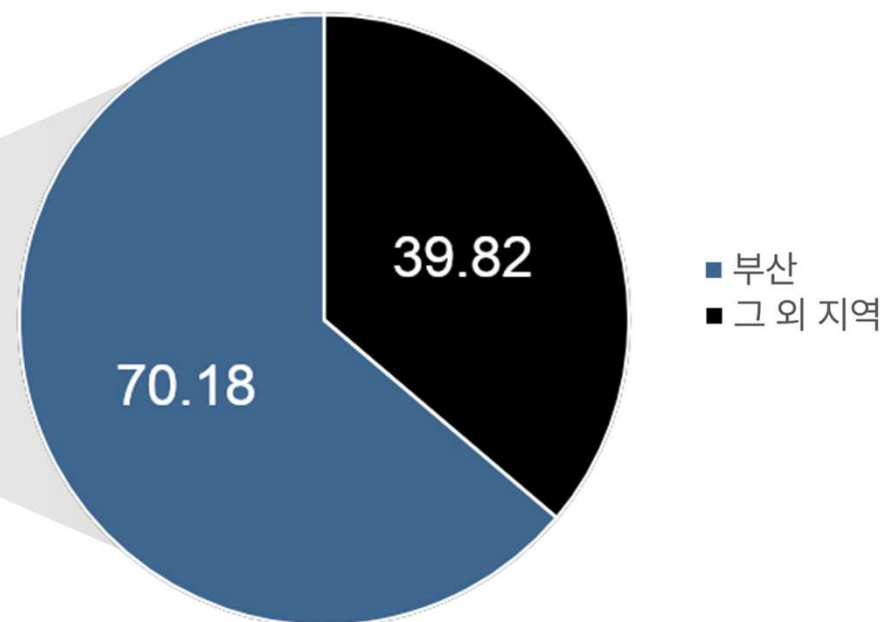
가설설계

:출산율 감소로 인해 학령인구 감소로 부산 사립 대학 존폐에 영향을 미칠 것이다.

[부산 내 고등학교 졸업생
부산 소재 대학교 입학 비율]

연도	2015	2016	2017	2018	2019
비율(%)	71.3	71.3	69.3	688.8	70.2

[5년 평균치 (단위:%)]



부산 내 고등학생의 부산 소재 대학교 입학 비율은 2015~2019년 평균 70.18%으로 확인되었음

Part 2

부산 인구 분포도 분석 방법 : 2015~2023 년 연령_및_성별_인구_읍면동

총 인구수 추출

```
dlsrn1 = dlsrn[dlsrn['항목'] == '총인구(명)']
dlsrn1
```

구	읍면동	항목	2015 년	2016 년	2017 년	2018 년	2019 년	2020 년	2021 년	2022 년	2023 년
0	부산광역시	합계	3440737.0	3440484.0	3416918.0	3395278.0	3372692.0	3349016.0	3324335.0	3295760.0	3279604.0
8	부산광역시	0~4세	135233.0	131709.0	124080.0	117270.0	106650.0	98637.0	89025.0	81941.0	75875.0
16	부산광역시	5~9세	128013.0	131323.0	131344.0	130641.0	131200.0	131903.0	129223.0	122846.0	116560.0
24	부산광역시	10~14세	142028.0	132615.0	129597.0	127020.0	124504.0	126187.0	129029.0	130198.0	129760.0
32	부산광역시	15~19세	204961.0	194712.0	182238.0	170236.0	159525.0	144199.0	133751.0	131316.0	131007.0
40	부산광역시	20~24세	248602.0	245825.0	240049.0	234572.0	226140.0	218047.0	208829.0	194156.0	184866.0
48	부산광역시	25~29세	212741.0	213370.0	218670.0	223464.0	225555.0	226643.0	225547.0	218324.0	210961.0
56	부산광역시	30~34세	239683.0	234194.0	208550.0	197580.0	192974.0	192718.0	195096.0	201175.0	206459.0
64	부산광역시	35~39세	245563.0	253126.0	255424.0	253798.0	242964.0	227062.0	212116.0	198745.0	190344.0
72	부산광역시	40~44세	268002.0	254784.0	244894.0	236317.0	232814.0	236830.0	243155.0	246962.0	247355.0
80	부산광역시	45~49세	282139.0	284928.0	282671.0	276778.0	270887.0	260000.0	246880.0	238052.0	230346.0
88	부산광역시	50~54세	295246.0	279576.0	270775.0	269089.0	272717.0	273734.0	275962.0	274172.0	269435.0
96	부산광역시	55~59세	308539.0	315908.0	313603.0	308391.0	296549.0	281623.0	266432.0	259060.0	259256.0
104	부산광역시	60~64세	239182.0	254676.0	264249.0	276347.0	285494.0	291543.0	299182.0	296958.0	293765.0
112	부산광역시	65~69세	179266.0	187551.0	195683.0	198815.0	206460.0	224086.0	238468.0	248079.0	260130.0
120	부산광역시	70~74세	139102.0	137409.0	140072.0	146932.0	155901.0	164547.0	172378.0	180497.0	184413.0
128	부산광역시	75~79세	97349.0	104388.0	113527.0	119287.0	121817.0	122339.0	121326.0	124066.0	130956.0
136	부산광역시	80~84세	53120.0	58478.0	62714.0	67136.0	72969.0	77700.0	83456.0	91218.0	96007.0
144	부산광역시	85~89세	23420.0	26094.0	27890.0	29656.0	32198.0	35683.0	38874.0	41440.0	44390.0
152	부산광역시	90~94세	7123.0	8080.0	8855.0	9707.0	10948.0	11847.0	12756.0	13434.0	14129.0
160	부산광역시	95~99세	1264.0	1560.0	1785.0	2063.0	2153.0	2384.0	2513.0	2722.0	3080.0
168	부산광역시	100세 이상	161.0	178.0	208.0	219.0	273.0	304.0	337.0	369.0	396.0
176	부산광역시	평균연령	42.0	42.6	43.1	43.6	44.2	44.8	45.3	45.9	46.4

부산광역시 연령 및 성별 전체
데이터셋에서 총 인구수 행 추출

연령별 분류

```
# 데이터프레임에서 필요한 열 추출
dlsrn2_grouped = pd.DataFrame([
    '연령대': ['미성년', '청년', '중년', '장년', '노년']
])

# 연도별로 계산
years = ['2015 년', '2016 년', '2017 년', '2018 년', '2019 년', '2020 년', '2021 년', '2022 년', '2023 년']
for year in years:
    dlsrn2_grouped[year] = [
        dlsrn2.loc[8:32, year].sum(), # 미성년 : 0~18
        dlsrn2.loc[40:56, year].sum(), # 청년 : 19~34세,
        dlsrn2.loc[64:80, year].sum(), # 중년 : 35~49세,
        dlsrn2.loc[88:104, year].sum(), # 장년 : 50~64세,
        dlsrn2.loc[112:168, year].sum(), # 노년 : 65세 이상
    ]

dlsrn2_grouped
```

연령대	2015 년	2016 년	2017 년	2018 년	2019 년	2020 년	2021 년	2022 년	2023 년
0 미성년	608235.0	590359.0	567259.0	545167.0	523879.0	500926.0	481028.0	466301.0	453216.0
1 청년	701026.0	683389.0	667269.0	655616.0	644669.0	637608.0	629472.0	613655.0	602286.0
2 중년	795704.0	792838.0	783029.0	766893.0	746665.0	724692.0	702151.0	683759.0	667945.0
3 장년	842967.0	850160.0	848627.0	853807.0	854760.0	846900.0	841576.0	830190.0	822456.0
4 노년	500805.0	523738.0	550734.0	573795.0	602719.0	638890.0	670108.0	701855.0	733701.0

미성년 : 0~18, 청년: 19~34세, 중년: 35~49세,
장년: 50~64세, 노년: 65세 이상으로 분류

Part 2

부산 사립대 지원자 수 분석 방법

: 2014 – 2024 년 고등 학교별X학과별 입학정원 지원 입학 학생 외국인학생 졸업 교원

4년제 대학 추출 및 폐교

```
[ ] volunteer = [] # 지원자 수
number_of_people_recruited = [] # 모집인원
for i in range( len(std) ):
    df = std[i]
    df = df_copy(df) # nan값 과 columns 재배열
    df = years_4_filtering(df) # 4년제 대학 추출
    df = df[df['시도'] == '부산' ]
    df = df[df['설립'] == '사립']
    df = df[df['학교상태'] != '폐교']
    volunteer.append(df['지원자_전체_계'].sum())
    number_of_people_recruited.append(df['모집인원_학부_계'].sum())
```

```
def years_4_filtering(df):
    d = df[df['학제'].isin( ['대학교', '산업대학', '교육대학'])]
    return d
```

Std[i]의 전체 데이터 프레임을 가져와 Years_4_filtering(df) 함수를 통해 4년제 대학교를 추출하고, 폐교를 제외한 ‘부산’, ‘사립’의 값을 가진 전체 지원자 수 추출

Part 2

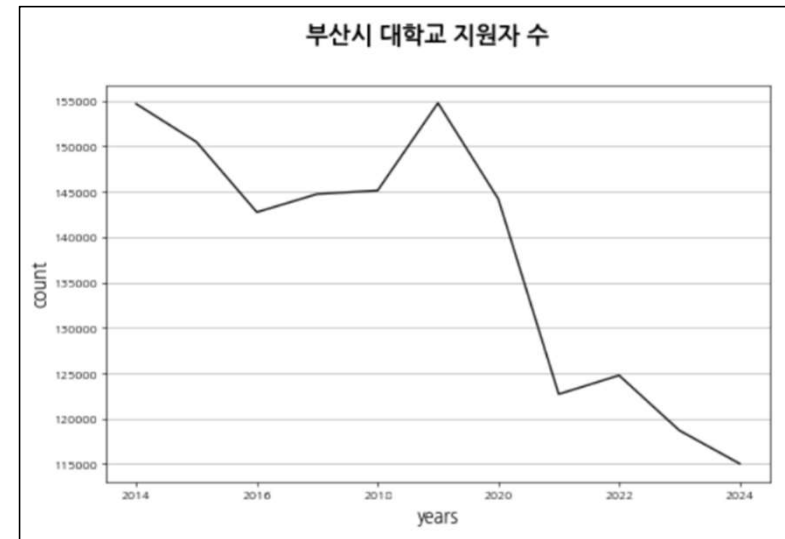
가설검증

: 출산율 감소로 인해 학령인구 감소로 부산 사립 대학 존폐에 영향을 미칠 것이다.

[2015~2023]



[2014~2024]



부산 인구 중 **미성년과 청년의 수**와 **부산 사립 대학교 지원자 수**의 지속적인 감소로 이 둘의 상관관계가 있음을 확인하여 인구 감소로 학령인구 감소로 부산 사립 대학의 존폐에 영향을 미친다는 가설을 확인함

- 1 2014~2024년도의 11년치 데이터의 방대한 양을 효율적으로 작업
- 2 Dataset의 이상치 및 결측치 제거
- 3 Dataset 원본손상 우려로 작업하기 전, `copy()` 함수 사용
- 4 실질적인 분석에 사용하려는 작업은 함수화 하여 공동작업에 효율성을 높임

Part 2

Dataset preprocessing : EDA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	S	T	U
1	<div> <div>■ 2014년 고등교육통계 학교별 학과별 주요 현황</div> <div>1. 셀 병합으로 생긴 cell에 NaN값</div> </div>																
2																	
3	1) 조사 기준일: 매년 4. 1. / 자료 추출일: 2022.10.18.																
4	* 입학자: 당해연도 3월 입학자 기준, 졸업자: 당해연도 2월 및 전년도 8월 졸업자 기준																
5	2) 요약정보 sheet에 학과수, 입학정원, 입학자, 재적학생수, 졸업자수, 전임교원수 피벗테이블이 있음																
6	3) 각 항목별 정의는 용어 정의 sheet 참조																
7	4) 학과 계열 구분은 「학과(전공)분류자료집」(한국교육개발원)에 따름																
8	* 출처 : 한국교육개발원 교육통계센터(국가교육통계센터) 교육통계 데이터베이스 (https://kess.kedi.re.kr)																
9	* 본 통계는 교육기본통계조사 결과로, 조사기준일에 따른 통계이며 공표 후 변동 없음. 유사 자료인 정보공시 자료는 수시로 수정 가능한 공시 자료로 본 통계와 다를 수 있음																
10																	
11	2. 실제로 사용할 columns는 12번째 존재. 기본 정보																
12																	
13																	
14	연도	학제	대학원구분	학교명	학교상태	본교	시도	시군구	설립	주야구분	주소	우편번호	전화번호	팩스번호	홈페이지	학과코드	학과명
314	2014	대학교		강원대학교	기존	본교	강원	강원 춘천시	국립	주간	강원 춘천시 강원대학	200-70	033-250-6114	033-251-9556	http://www.gw.ac.kr	U07020300105	영상문화학과
315	2014	대학교		강원대학교	기존	본교	강원	강원 춘천시	국립	주간	강원 춘천시 강원대학	200-70	033-250-6114	033-251-9556	http://www.gw.ac.kr	U07030100005	무용학과
316	2014	대학교		강원대학교	기존	본교	강원	강원 춘천시	국립	주간	강원 춘천시 강원대학	200-70	033-250-6114	033-251-9556	http://www.gw.ac.kr	U07030200061	스포츠과학부
317	2014	대학교		강원대학교	기존	본교	강원	강원 춘천시	국립	주간	강원 춘천시 강원대학	200-70	033-250-6114	033-251-9556	http://www.gw.ac.kr	U07040100009	미술학과
318	2014	대학교		강원대학교	기존	본교	강원	강원 춘천시	국립	주간	강원 춘천시 강원대학	200-70	033-250-6114	033-251-9556	http://www.gw.ac.kr	U07060100003	음악학과
319	2014	대학교		강원대학교	기존	본교	강원	강원 춘천시	국립	주간	강원 춘천시 강원대학	200-70	033-250-6114	033-251-9556	http://www.gw.ac.kr	Z999999999999	소속학과없음
320	2014	대학교		강원대학교	기존	본교	강원	강원 춘천시	국립	주간	강원 춘천시 강원대학	200-70	033-250-6114	033-251-9556	http://www.gw.ac.kr		
321	2014	대학교		강원대학교	기존	제2캠퍼스	강원	강원 삼척시	국립	주간	강원 삼척시 중앙로 32	245-93	033-570-6114	033-572-8620	http://www.gw.ac.kr	U01010300009	일본어학전공
322	2014	대학교		강원대학교	기존	제2캠퍼스	강원	강원 삼척시	국립	주간	강원 삼척시 중앙로 32	245-93	033-570-6114	033-572-8620	http://www.gw.ac.kr	U01010600016	영미언어문화전공
323	2014	대학교		강원대학교	기존	제2캠퍼스	강원	강원 삼척시	국립	주간	강원 삼척시 중앙로 32	245-93	033-570-6114	033-572-8620	http://www.gw.ac.kr	U01010600017	영어과
324	2014	대학교		강원대학교	기존	제2캠퍼스	강원	강원 삼척시	국립	주간	강원 삼척시 중앙로 32	245-93	033-570-6114	033-572-8620	http://www.gw.ac.kr	U01010600092	영미언어문화학과
325	2014	대학교		강원대학교	기존	제2캠퍼스	강원	강원 삼척시	국립	주간	강원 삼척시 중앙로 32	245-93	033-570-6114	033-572-8620	http://www.gw.ac.kr	U01011200016	어문학부
326	2014	대학교		강원대학교	기존	제2캠퍼스	강원	강원 삼척시	국립	주간	강원 삼척시 중앙로 32	245-93	033-570-6114	033-572-8620	http://www.gw.ac.kr		
327	2014	대학교		강원대학교	기존	제2캠퍼스	강원	강원 삼척시	국립	주간	강원 삼척시 중앙로 32	245-93	033-570-6114	033-572-8620	http://www.gw.ac.kr		

3. 학교상태 검증

4. 일부 학과 코드와 학과명은 이상치값으로 존재

Part 2

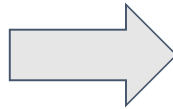
Dataset preprocessing

: std라는 하나의 변수에 list 자료형에 연도별 DataFrame 저장

```
# directory 저장.  
dir = '/content/drive/MyDrive/Data/고등 학교별X학과별 입학정원 지원 입학 학생 외국인학생 졸업 교원/'  
  
# os.listdir: 해당 directory의 file을 list로 읽어오는 코드  
# sorted() 정렬함수.  
folders = sorted( os.listdir(dir) )  
  
std = [] # 불러올 excel file을 저장할, list  
for i in folders:  
    # append()를 통해 해당 std 리스트에 excel 파일을 순서대로 삽입함.  
    std.append( pd.read_excel(dir + i) )
```

std[0] = 2014s
std[1] = 2015s
std[2] = 2016s

·
·
·



for i in range(len(std)):

std[i]

for i in std:

i

Part 2

Dataset preprocessing

: 실제로 사용할 columns 값들은 13번째 row에 존재함

```
# df 복사본 생성 및 columns 재배열.  
def df_copy(df):  
    # df 복사본 생성.  
    df_ = df.copy()  
    # columns 재 지정.  
    df_.columns = np.array(df_.loc[12])  
    # 기존 excel에 병합되어 생긴 nan 제거.  
    df_ = df_.loc[13:]  
    return df_
```

Part 2

Dataset preprocessing

: 행이 약 3만개이상, 열이 129개로 구성된 Excel 파일을 11년치 사용으로 load에 시간이 오래 걸려서 필요한 42개의 columns만을 남김

```
using_columns = ['학교명', '시도', '설립', '대계열', '중계열', '소계열', '학과코드', '학과명',  
                 '학과수_전체', '입학정원_학부_계', '정원내_입학정원_학부', '정원외_모집정원_학부',  
                 '지원자_전체_계', '지원자_전체_남', '지원자_전체_여', '입학자_전체_계', '입학자_전체_남', '입학자_전체_여',  
                 '정원내_입학자_학부_계', '정원내_입학자_학부_남', '정원내_입학자_학부_여', '정원외_입학자_전체_계',  
                 '정원외_입학자_전체_남', '정원외_입학자_전체_여', '재적생_전체_계', '재적생_전체_남', '재적생_전체_여',  
                 '재학생_전체_계', '재학생_전체_남', '재학생_전체_여', '휴학생_전체_계', '휴학생_전체_남', '휴학생_전체_여',  
                 '외국인유학생_학사_계', '외국인유학생_학사_남', '외국인유학생_학사_여', '졸업자_전체', '졸업자_남', '졸업자_전체_여',  
                 '전임교원_계', '비전임교원_계', '시간강사_계']  
  
# df 복사본 생성 및 columns 재배열.  
def df_copy(df):  
    # df 복사본 생성.  
    df_ = df.copy()  
    # columns 재 지정.  
    df_.columns = np.array(df_.loc[12])  
    # 기존 excel에 병합되어 생긴 nan 제거.  
    df_ = df_.loc[13:]  
    df_ = df_[using_columns]  
    return df_
```

Part 2

Dataset preprocessing - Dataset에서 의미하는 폐교

사회 > 부산 / 경남

[부산/경남]동명정보대-동명대 통합 명칭'동명대학교'

업데이트 2009-10-08 21:15

🔍 📄 🗨️ 📧

🔍 📄 🗨️ 📧

2006년 3월 통합되는 동명정보대(4년제)와 동명대(2년제)의 이름이 동명대학교로 확정됐다.

학교법인 동명문화학원은 통합대학 명칭을 동명대학교(東明大學校·Tongmyong University)로 정했다고 26일 밝혔다.

이 대학은 9월 수시 2학기 모집부터 양 대학의 학과를 조정해 6개 단과대학(33개 학과)과 1개 학부(2개 전공) 체제로 신입생을 모집하고 기존 IT분야 특성화를 가속화하는 등 구조개혁을 추진하겠다고 밝혔다.

```
for i in range(len(std)):  
    df = df_copy(std[i])  
    print(f"{folders[i][:4]}: {df[df['학교명'] == '동명정보대학교']['학교상태'].unique()}")
```

2014: ['폐교']

2015: ['폐교']

2016: []

2017: []

2018: []

2019: []

2020: []

2021: []

2022: []

2023: []

2024: []

	재적생_전체_계	재학생_전체_남	휴학생_전체_계	졸업자_전체
학교명				
동명정보대학교	235	75	155	148
2015				
학교명				
동명정보대학교	0	0	0	50
2016				

Empty DataFrame

Columns: [재적생_전체_계, 재학생_전체_남, 휴학생_전체_계, 졸업자_전체]

Index: []

Part 2

Dataset preprocessing - Dataset에서 의미하는 폐교

	연도	학제	학교명	학교상태	학과명
23975	2014	산업대학	우송대학교	폐교	영어학과
23976	2014	산업대학	우송대학교	폐교	미국유학과
23977	2014	산업대학	우송대학교	폐교	일본학과
23978	2014	산업대학	우송대학교	폐교	중국학부
23979	2014	산업대학	우송대학교	폐교	국제경영학부
23980	2014	산업대학	우송대학교	폐교	호텔관광경영학과
	연도	학제	학교명	학교상태	학과명
24087	2015	산업대학	우송대학교	폐교	호텔관광경영학과
24088	2015	산업대학	우송대학교	폐교	건축학부
24089	2015	산업대학	우송대학교	폐교	철도건설환경공학과
24090	2015	산업대학	우송대학교	폐교	컴퓨터정보학과
24091	2015	산업대학	우송대학교	폐교	컴퓨터디자인학과
24092	2015	산업대학	우송대학교	폐교	스포츠건강관리학부

학교 폐교가 아닌,
학과 폐지를 의미

Part 2

Dataset preprocessing - 이상치 제거

학교명	학과코드 (고유명)	학과코드 (수)	학과명 (고유명)
가천의과학대학교	[Z99999999999]	1	[소속학과없음]
가톨릭대학교	[Z99999999999]	2	[소속학과없음]
가톨릭상지대학	[Z99999999999]	1	[소속학과없음]
감리교신학대학교	[Z99999999999]	1	[소속학과없음]
강릉대학교	[Z99999999999]	1	[소속학과없음]
...
대전대학교	[Z99999999999]	1	[소속학과없음]
대전대학교	[Z99999999999]	1	[소속학과없음]
호남대학교	[Z99999999999]	1	[소속학과없음]
호서대학교	[Z99999999999]	1	[소속학과없음]
웨슬리언대학교	[Z99999999999]	1	[소속학과없음]

378 rows × 3 columns

Part 2

Dataset preprocessing

- 학과의 분류체계 (대 분류: 5개, 중분류: 31개, 소분류: 약 120개 그리고 학과로 세분화

인문사회 계열 (6소계열)			자연과학 계열 (8소계열)			공학 계열 (7소계열)			예체능 계열 (6)				
언어 문학	언어학 국어국문학 독일어문학 러시아어문학 스페인어문학 영어영문학 일본어문학 중국어문학 프랑스어문학 기타아시아어문학 기타유럽어문학 교양어문학	사회 과학	수학 통계학 물리학 천문/기상학 지구과학 해양학 반도체과학	생활 과학	식품영양학 조리과학 의류/의상학 주거학 의예과 치위예과 한위예과 수의예과	건설	건축학 건축공학 조경학 토목공학 도시공학 환경공학	재료	금속공학 반도체공학 신소재공학 세라믹공학 재료공학 화학공학	무용 체육	무용 체육		
	화학 생명 환경		의로 예과	기계공학 기전공학 조선/해양공학 항공/우주공학 철도공학 자동차공학	에너지공학 고분자공학 생명공학 섬유공학 산업공학 안전공학 방재공학 공학교육		연극 영화		연극 영화 방송연예				
	농림 수산		약학	전기공학 전자공학 제어계측공학	화공 고분자 에너지							미술	공예 디자인 순수미술 응용미술 미술학
인문	철학/윤리학 역사/고고학 종교학 문화/민속/ 미술사학 국제지역학 교양인문학	경영 경제	작물/원예학 산림학 축산학 수산학 농림수산환경 생태학 농림수산바이오 시스템공학 식품가공학	보건	보건학 재활치료 임상보건 보건관리 피부미용	전기 전자 컴퓨터		전기공학 전자공학 제어계측공학		산업 안전	산업공학 안전공학 방재공학 공학교육 교양공학		
	교육학 중등언어교육 중등사회교육 유아교육학 초등교육 특수교육학		교육	중등자연 과학교육 간호/보건교육 교양자연과학	응용소프트 웨어공학 의공학 광학공학		교육	음악이론					
	의약 계열		의학 치위학 한위학 수의학	전산학/ 컴퓨터공학 정보/통신공학	응용 예술				사진 만화				
교육	교육학 중등언어교육 중등사회교육 유아교육학 초등교육 특수교육학	법학	의로	의학 치위학 한위학 수의학		기타				응용 예술	사진 만화	응용 예술	애니메이션 게임 영상예술 음향
	광고/홍보학 관광학 부동산 법학						기타	비서학					
	법학				비서학				뷰티아트				

Part 2

Dataset preprocessing

: 학과의 분류체계

(대계열: 5개, 중계열: 31개, 소계열: 약 120개 그리고 학과로 세분화)

학교명	대계열	중계열	소계열	학과명
경성대학교	공학계열	건축	건축학	건축학부

학과 개편

- 2015(건축학과)

- 2017(디자인 건축학부)

Part 2

Dataset preprocessing

: 소계열로 그룹화 하여, 경쟁률 계산

```
result = []
for i in range( len(std) ):
    df = df_copy(std[i])
    university = df['학교명'].unique().tolist()

    for j in range( len(university) ):
        df_ = df[df['학교명'] == university[j]]
        df_ = df_.groupby('소계열')['학과명'].count().reset_index()
        df_['University'] = university[j]
        df_['Year'] = df['연도'].loc[:]
        df_ = df_[['Year', 'University', '소계열', '학과명']]

        df_subset = df.groupby('소계열')[['입학자_전체_계', '정원내_입학정원_학부', '지원자_전체_계']].sum()
        df_ = pd.merge(df_, df_subset, on='소계열', how='left')

    result.append(df_)
```


Part 2

Dataset preprocessing

: 고등교육기관은 9가지가 존재하며, 전문성에 따라 19개로 세분화 되어 있었음

과정	교육기관
전문학사학위	전문대학
	사이버대학
	기술대학
	각종학교
	전공대학
	사내대학
	원격대학
	기능대학
	한국농수산대학

```
def years_4_filtering(df):  
    d = df[df['학제'].isin(['대학교', '산업대학', '교육대학'])]  
    return d
```

```
std_['학제'].unique()
```

```
array(['대학교', '일반대학원', '특수대학원', '전문대학원', '교육대학', '산업대학', '방송통신대학교',  
      '기술대학', '사내대학(전문)', '사내대학(대학)', '전문대학', '각종대학(전문)', '각종대학(대학)',  
      '원격대학(전문)', '원격대학(대학)'], dtype=object)
```

Part 2

Dataset preprocessing

: 비율 계산시 문제되는 0값을 1e-9 대체함.

```
-----  
-----  
ZeroDivisionError                                Traceback (most recent call last)  
<ipython-input-147-39ba9e6117ee> in <cell line: 0> ()  
    16 # 유학생 비율 계산 (유학생 수 / 재적자수) - ZeroDivisionError 해결  
    17 result['유학생비율'] = np.where(result['전체_외국인유학생'] != 0,  
--> 18                                result['전체_외국인유학생'] / result['전체_재적학생수'] * 100, 0)  
    19  
    20 # 출력  
  
-----  
↕ 8 frames  
/usr/local/lib/python3.11/dist-packages/pandas/core/computation/expressions.py in _evaluate_standard  
    71 if _TEST_MODE:  
    72     _store_test_result(False)  
--> 73 return op(a, b)  
    74  
    75  
ZeroDivisionError: division by zero
```

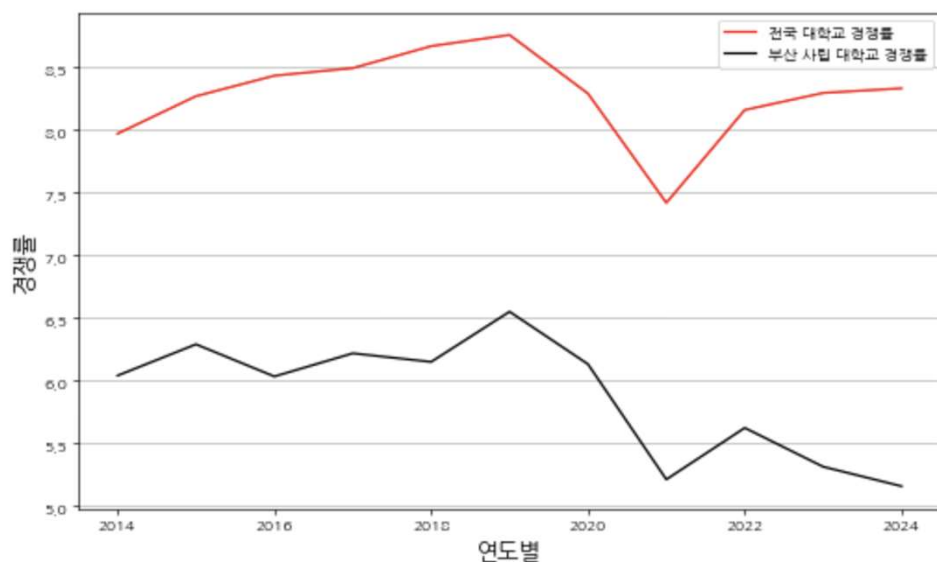
오류 발생: 0값 존재.

A 대학 유학생 수

전체 대학교 유학생 수

“부산에 있는 대학교는 유학생 유치를 통해 존속을 한다.”

전국 대학교와 부산 사립대 경쟁률



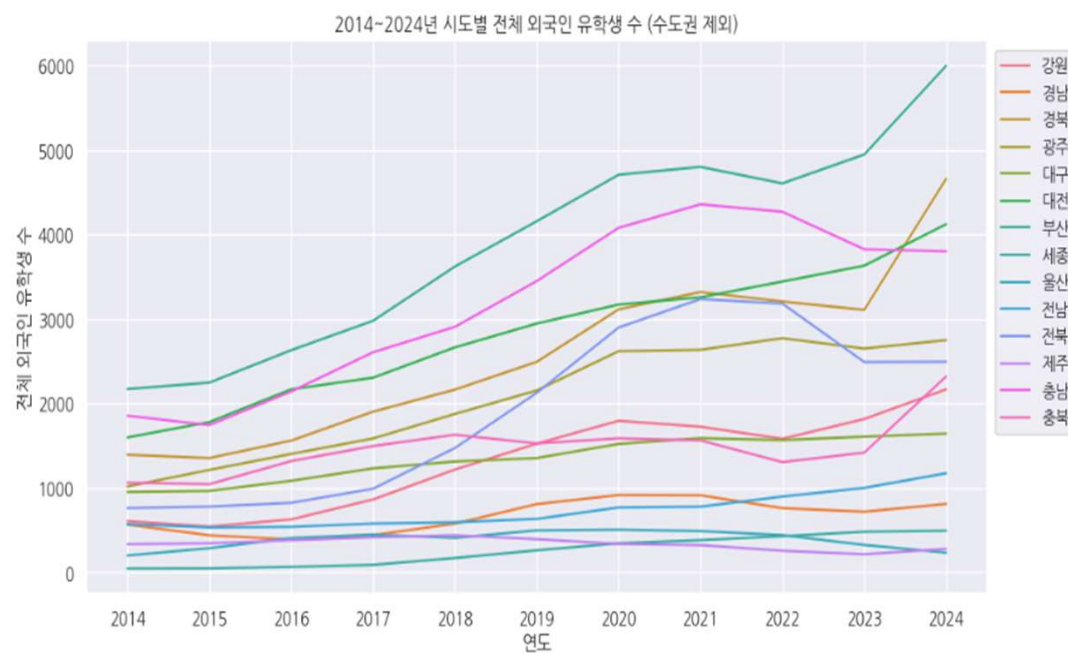
대학 폐쇄/자진폐교 현황

유형	구분	교명	재단명	폐교 시기	폐교 사유
폐쇄 명령 (7개교)	4년제	국제문화대학원대학교	국제문화대학원	2014년2월	- 부당학위/부당학점 부여 - 비정상적 수업 운영 적발 - 시정 요구 지속적 미이행
		선교청대	대정학원	2012년8월	- 학사운영/재정회계/교직원임용 등 위법/부당 적발 - 시정요구 지속적 미이행
		명신대	신명학원	2012년2월	- 수익용 기본재산 무단처분 - 학생성적관리/교수채용 등 부정적 적발 - 시정요구 지속적 미이행
		아시아대	아시아교육재단	2008년2월	- 허위 재산출연 교수 - 직원 채용시 금품수수 등 비리 적발 - 2005년 감사 결과 시정 요구사항 미이행
	전문대	벽성대	충렬학원	2014년8월	- 부당학위/부당학점 부여 - 비정상적 수업 운영 적발 - 시정요구 지속적 미이행, 재정부실로 정상적 운영 불가능
		성화대	세림학원	2012년2월	- 회계부정/부당학위 등 학사운영 부실 적발 - 시정요구 지속적 미이행
	각종학교	개혁신학교	개혁신학원	2008년2월	- 임원간 분쟁 등으로 정상적인 학사운영 불가능 - 부채해결/학교정상화 방안 제출요구 불이행
자진 폐쇄 (3개교)	4년제	경북외대	경북외국어대	2014년2월	- 학생충원난, 인건비 상승 등 재정적 어려움
		건동대	백암교육재단	2013년2월	- 학생충원난으로 재정적 어려움 - 부당학위/부당학점 부여
		광주예대	하남학원	2000년2월	- 설립자 비리, 대학 부실운영적발 등 정상적인 운영 불가능

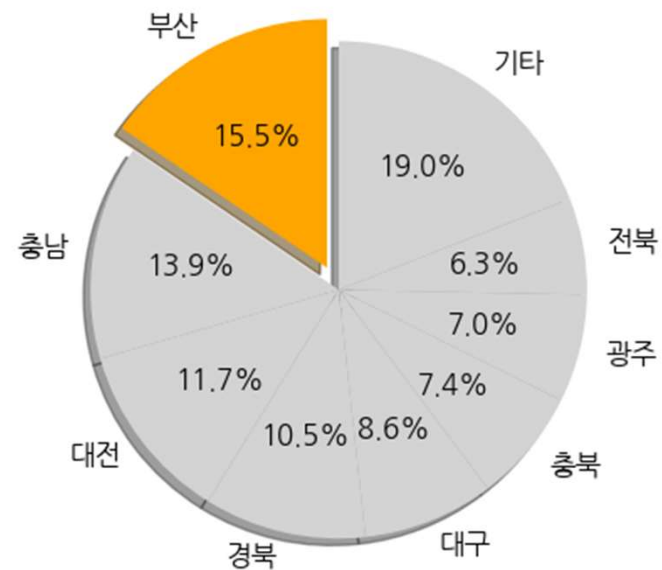
*출처=교육부

Part 3

전국 유학생 추이

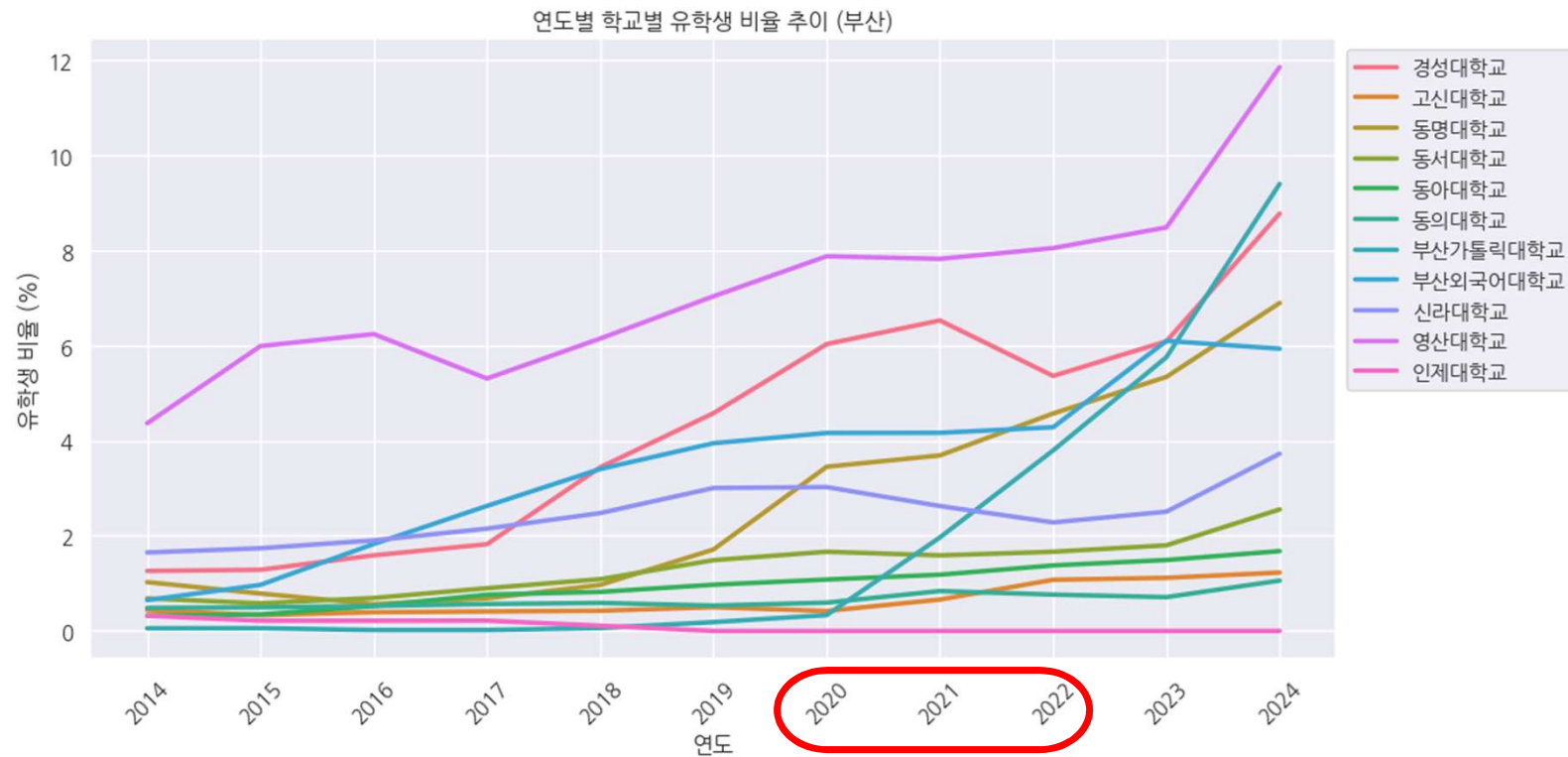


014년도 시도별 전체 외국인 유학생 비율 (수도권 제외, 5% 미만 기타)



Part 3

코로나시기 유학생 추이 분석

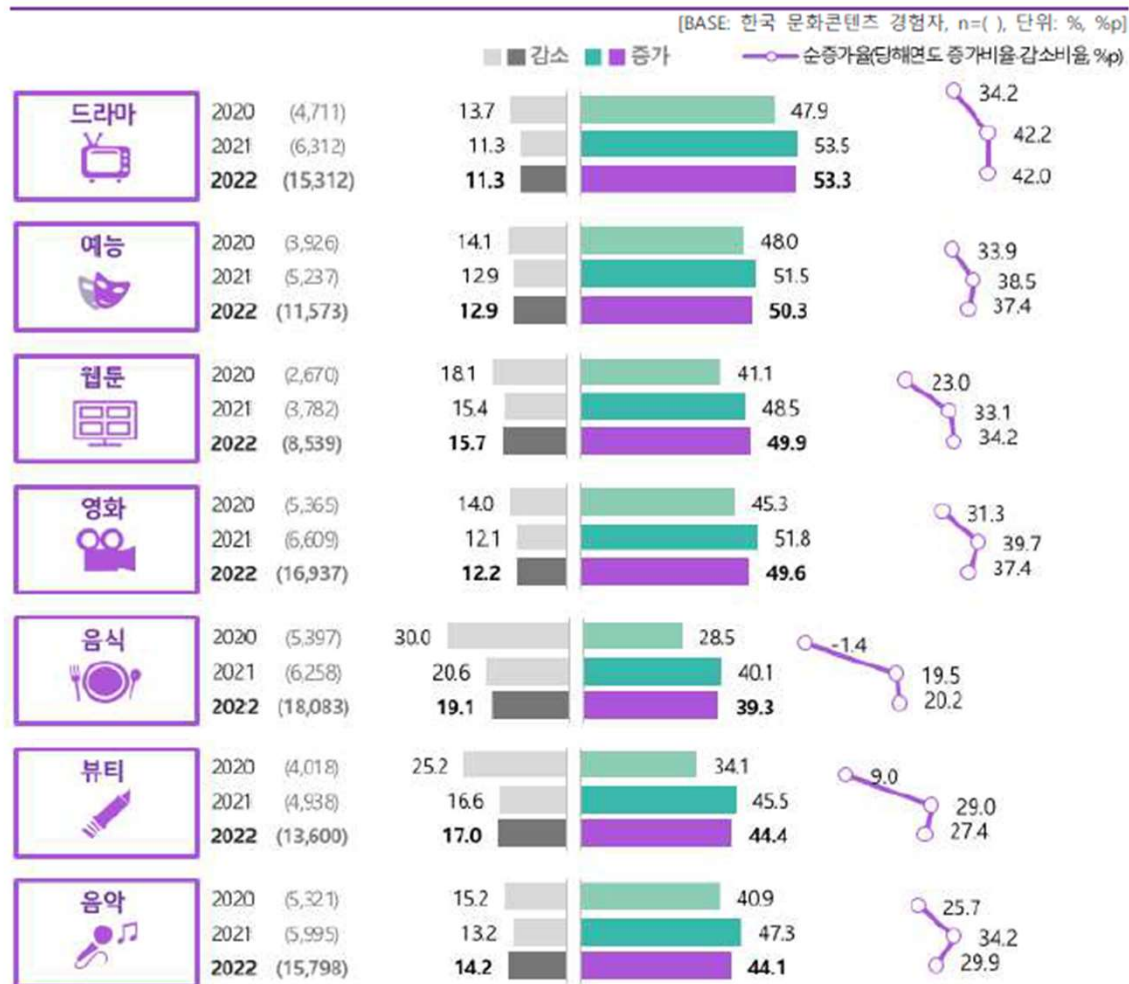


“코로나임에도 유학생 유치할 수 있었다는 것은 주목할 만한 지표이다.”

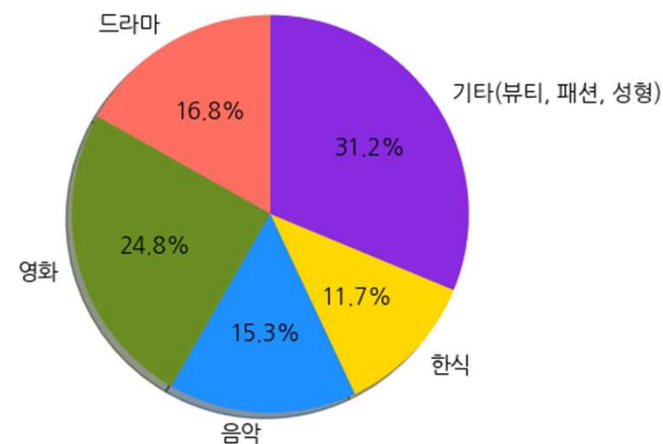
Part 3

해외 이용자의 한국 콘텐츠 소비량 탐색

[그림 2-4] 코로나19 발생 이전 대비 한국 문화콘텐츠 소비 변화추이



2020년도 연령별 한국 콘텐츠 선호도



출처: 한국국제문화교류진흥원

Part 3

해외 이용자의 한국 콘텐츠 소비량 탐색

[표 2-3] 최선호 한국 문화콘텐츠 변화추이

[BASE: 한국 문화콘텐츠 경험자, n=(), 단위: 1순위%]

		2018년	2019년	2020년	2021년	2022년
드라마	사례수	(5,235)	(5,563)	(4,143)	(4,850)	(11,647)
	1위	미스터 션샤인 4.7	호텔 델루나 8.2	사랑의 불시착 9.5	오징어 게임 21.2	오징어 게임 11.3
	2위	태양의 후예 3.7	태양의 후예 3.0	사이코지만 괜찮아 4.1	사랑의 불시착 2.2	이상한 변호사 우영우 2.8
	3위	도깨비 3.4	대장금 2.3	부부의 세계 2.8	빈센조 1.9	지금 우리 학교는 2.6
	4위	김비서가 왜 그럴까 2.9	킹덤 2.1	이태원 클라쓰 2.8	펜트하우스 1.6	사내맞선 1.9
	5위	너도 인간이니? 2.6	도깨비 2.0	킹덤 2.5	갯마을 차차차 1.4	사랑의 불시착 1.9
영화	사례수	(5,482)	(5,801)	(4,704)	(5,048)	(12,837)
	1위	부산행 8.4	기생충 12.4	기생충 18.4	기생충 10.3	기생충 9.1
	2위	지금, 만나고 갑니다 5.8	부산행 7.9	부산행 10.2	부산행 6.8	부산행 6.9
	3위	신과함께: 죄와 벌 4.5	극한직업 3.1	반도 3.5	서복 1.5	카티 2.5
	4위	신과함께: 인과 연 2.1	신과함께 2.0	#살아있다 2.1	미드나이트 1.3	헌트 2.0
	5위	버닝 2.0	엑시트 1.5	사냥의 시간 1.6	승리호 1.3	20세기 소녀 1.8
애니메이션	사례수	(5,740)	(4,252)	(3,782)	(4,056)	(10,174)
	1위	루카 20.1	루카 15.9	루카 15.0	라바 9.6	루카 7.7
	2위	쵸로로 19.8	쵸로로 15.5	라바 11.8	쵸로로 9.3	신비아파트 7.5
	3위	라바 16.9	라바 14.8	쵸로로 11.6	루카 8.8	라바 7.4
	4위	타요 11.6	핑크퐁 13.3	핑크퐁 9.5	신비아파트 8.0	서울역 6.5
	5위	폴리 10.5	슈퍼원즈 10.3	슈퍼원즈 7.1	슈퍼원즈 7.3	쵸로로 6.2
게임	사례수	(3,859)	(3,746)	(3,507)	(3,687)	(9,856)
	1위	배틀그라운드 18.0	배틀그라운드 18.3	배틀그라운드 18.0	배틀그라운드 14.5	배틀그라운드 10.3
	2위	라그나로크 15.7	라그나로크 11.4	라그나로크 11.6	라그나로크 12.1	라그나로크 9.7
	3위	크로스파이어 11.0	크로스파이어 10.5	크로스파이어 9.8	크로스파이어 9.7	마블 퓨저 레볼루션 9.1
	4위	검은사막 8.8	검은사막 8.9	검은사막 7.5	검은사막 6.6	크로스파이어 7.5
	5위	메이플스토리 8.6	던전앤파이터 8.5	서머너즈 워 7.2	던전앤파이터 6.2	검은사막 5.9

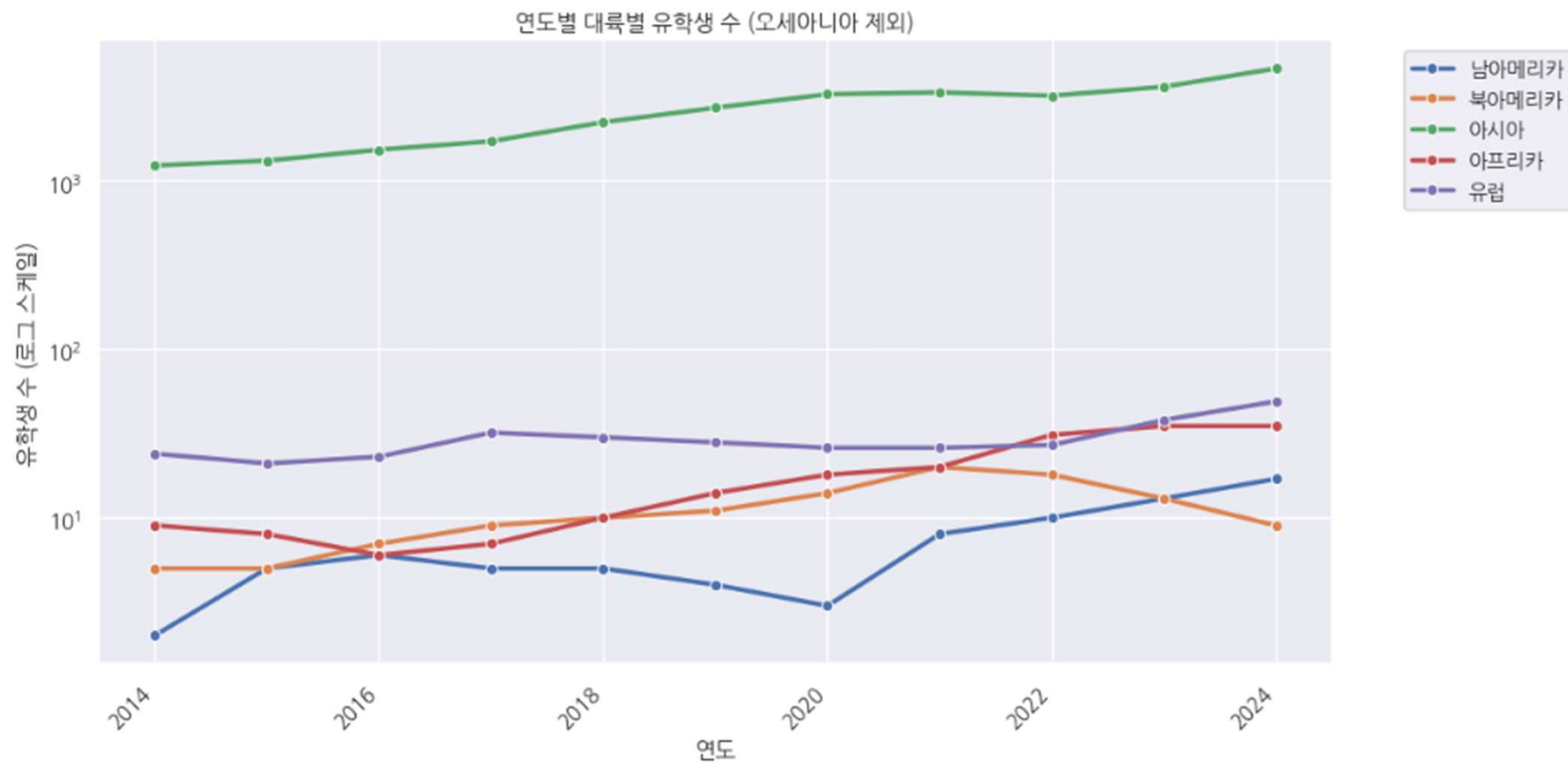
[표 2-4] 최선호 한류스타 변화추이

[BASE: 전체(배우)/한국 문화콘텐츠 경험자(가수/그룹), n=(), 단위: 1순위%]

		2018년	2019년	2020년	2021년	2022년
배우	사례수	(7,500)	(8,000)	(8,500)	(8,500)	(25,000)
	1위	이민호 8.6	이민호 7.6	이민호 9.6	이민호 9.3	이민호 9.1
	2위	송혜교 2.4	송혜교 3.4	현빈 3.5	현빈 3.1	공유 2.7
	3위	공유 2.4	송중기 1.9	공유 2.3	공유 2.8	현빈 2.4
	4위	송중기 2.1	공유 1.8	송혜교 2.1	송혜교 2.2	송혜교 2.0
	5위	이종석 1.9	배수지 1.7	이종석 2.0	송중기 2.1	이종석 1.5
가수/그룹	사례수	(5,248)	(5,665)	(4,695)	(4,678)	(12,511)
	1위	방탄소년단 10.3	방탄소년단 15.2	방탄소년단 22.0	방탄소년단 26.7	방탄소년단 31.3
	2위	싸이 6.9	블랙핑크 6.5	블랙핑크 13.5	블랙핑크 10.4	블랙핑크 10.8
	3위	블랙핑크 3.8	싸이 5.7	싸이 2.9	아이유 2.8	싸이 2.6
	4위	빅뱅 2.5	트와이스 2.8	트와이스 2.4	리사 2.4	트와이스 1.8
	5위	엑소 2.3	빅뱅 2.5	엑소 2.1	싸이 2.2	아이유 1.7

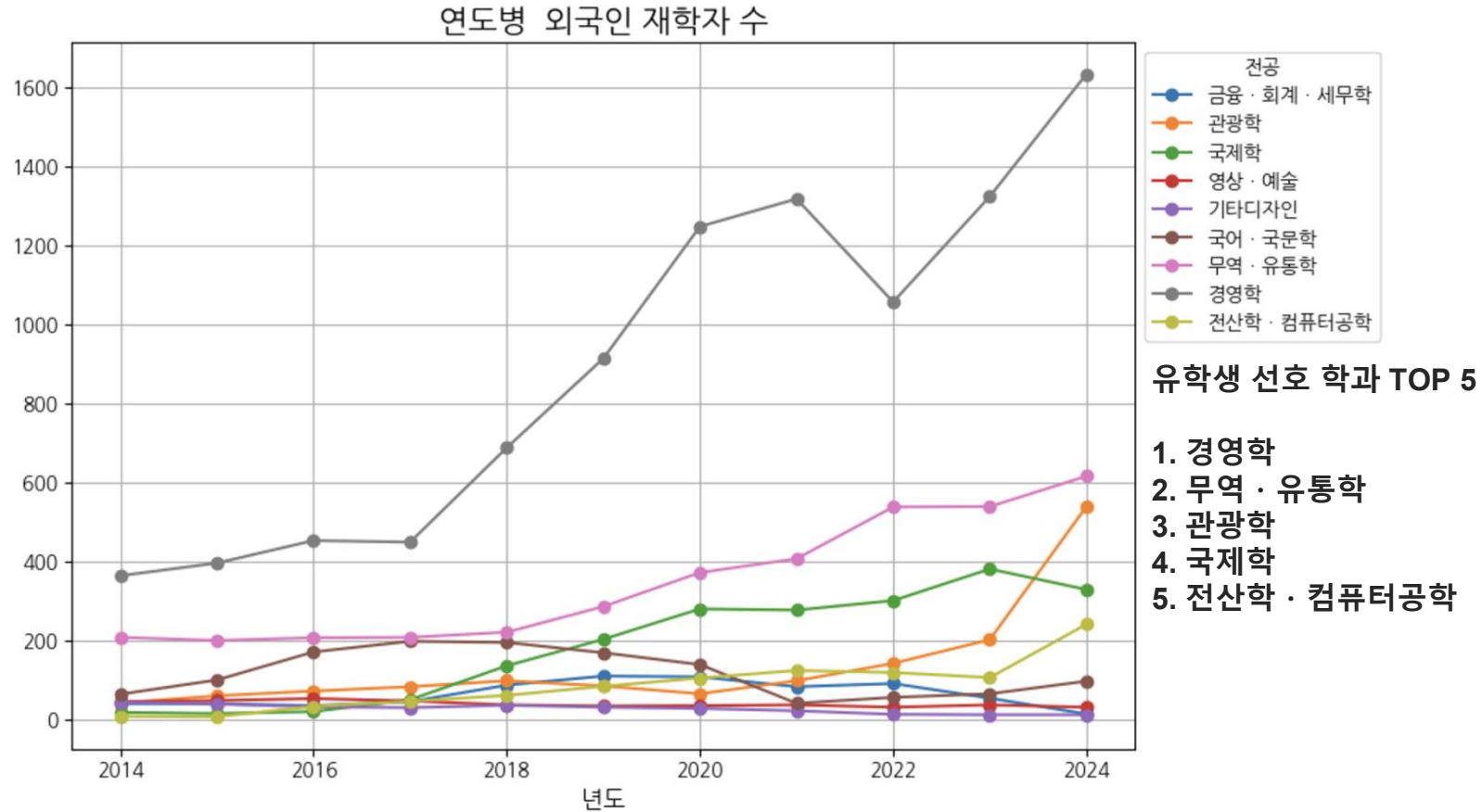
Part 3

부산 내 대학교 국가별 유학생 비율



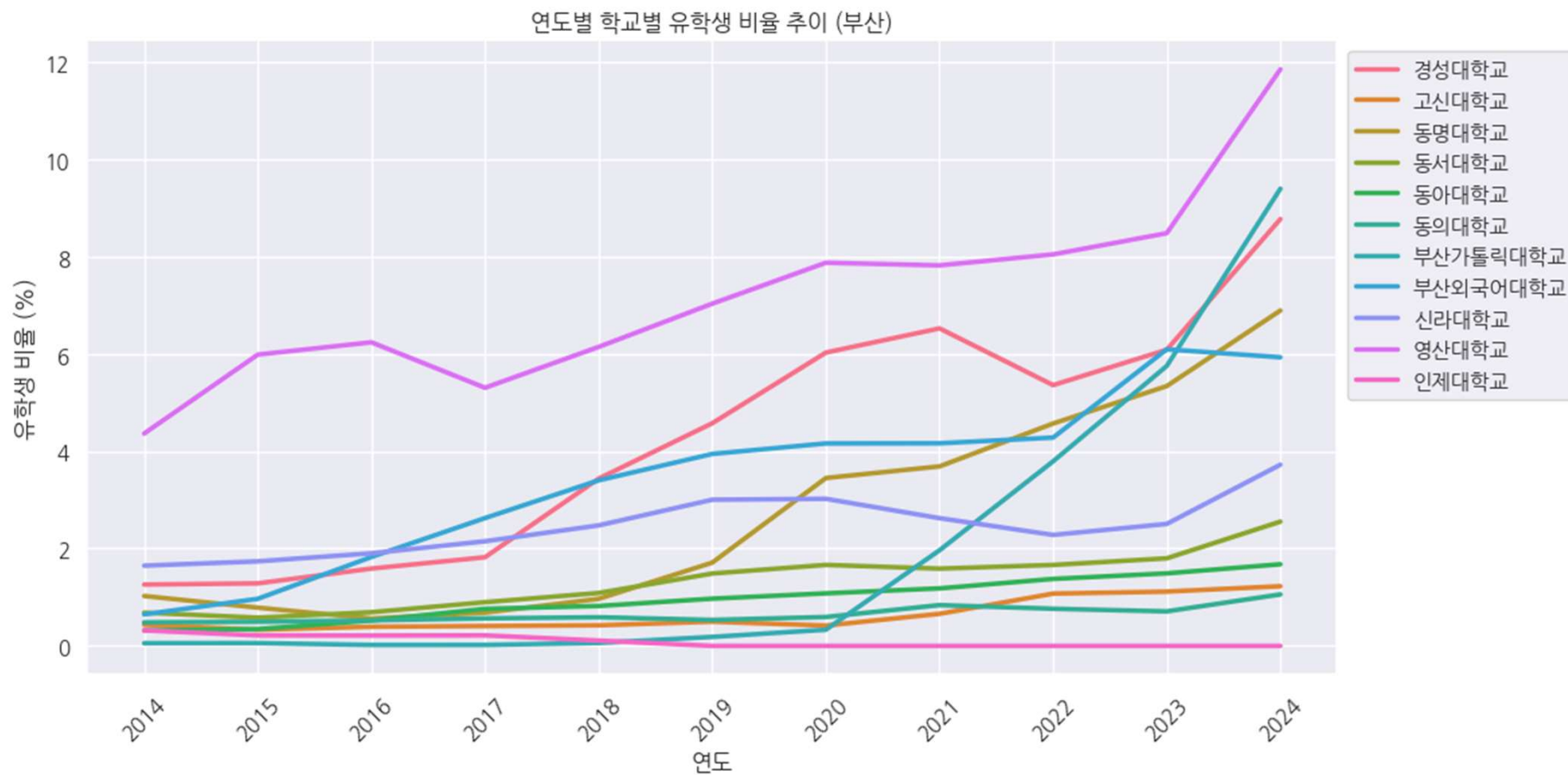
Part 3

부산 내 대학교 유학생 학과 선호도



Part 3

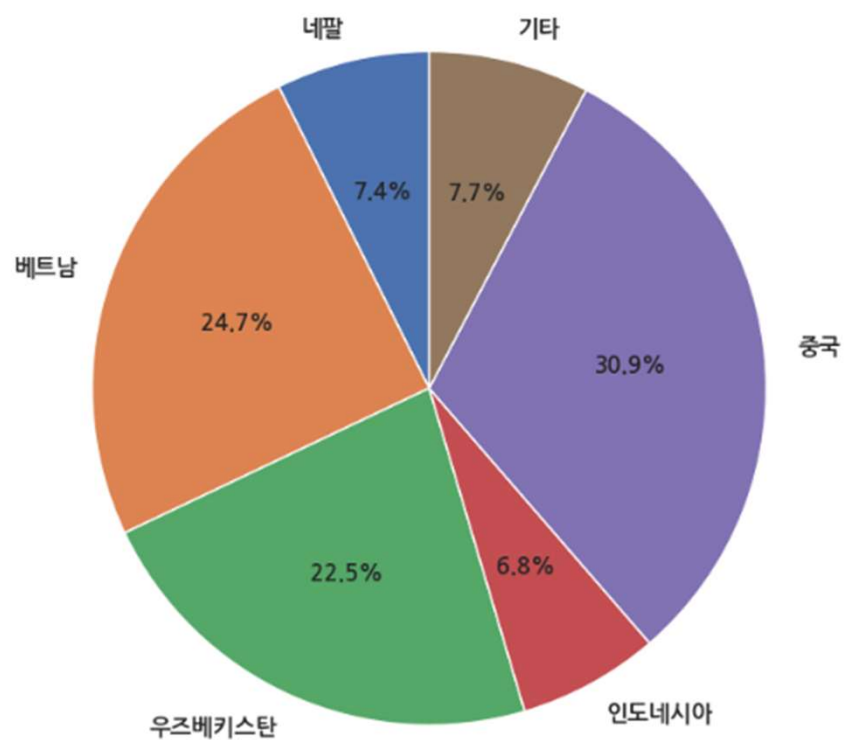
영산대학교 사례



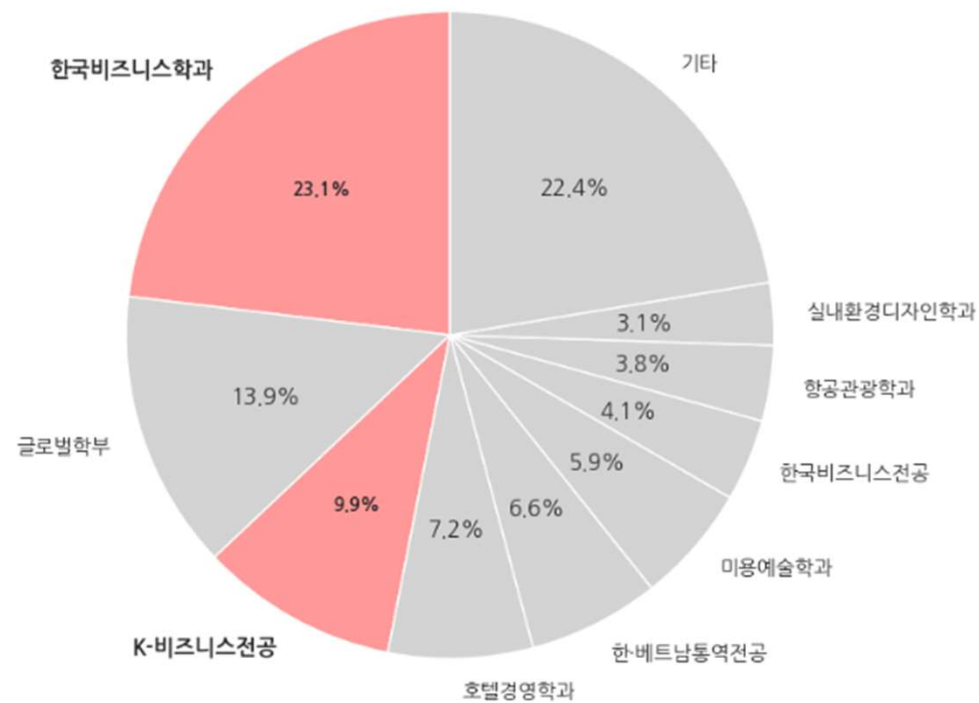
Part 3

영산대학교 사례

영산대학교 2014~2024년 유학생 국적 비율 (평균, 3% 미만 기타)



2014~2024년 영산대학교 유학생 학과별 분포 (평균, 3% 미만 기타)



**“ 부산의 4년제 사립대학교들은 학령인구 감소와 학생 선호도 하락으로 인해
국내 학생 수를 확보하는 데 어려움을 겪고 있습니다.**

**이러한 위기 속에서 한류 콘텐츠를 활용한 부산의 국제적 이미지 제고와
경영학 등 유망 학과의 개설 및 특화를 통해 유학생을 유치하는 전략이 중요합니다.**

**유학생 유치를 통해 대학들은 충원율 문제를 해결하고, 재정적 안정성을 확보함으로
써 현재의 위기를 극복하고 지속 가능한 발전을 이룰 수 있습니다. ”**

권수현

- 의견 조율의 어려움
 - 팀원들 간의 다양한 아이디어와 의견을 조율하는 과정에서 어려움을 겪음.
 - 모두가 만족할 수 있는 방향을 설정하기 위해 많은 논의가 필요했음.
- 데이터 분석 참여의 제한
 - 프로젝트 리딩과 아이디어 제공에 집중하며 전반적인 그림을 그리는데 집중하다보니, 데이터 분석 과정에 직접적으로 참여할 기회가 부족했던 점이 아쉬움.
 - 다양한 데이터를 검토하며 최적의 데이터를 선정하려 노력했지만, 초기 단계에서 난항을 겪음.
- 리더십 경험의 보람
 - 반면, 프로젝트 리딩 역할을 통해 팀을 이끌고, 큰 그림을 설계하며 협업의 중요성을 배울 수 있었음.
 - 이러한 경험이 개인의 리더십과 문제 해결 능력을 한 단계 성장시키는 계기가 됨.

김채영

예상과 다른 결과 분석의 어려움

- 부산 사립대학들의 미래를 유학생 유치를 통해 예측하는 과정에서, 연도별로 대륙, 국가, 학과, 학교별로 도출된 유학생 데이터가 예상과 다른 결과를 나타냄.
- 데이터의 방대함으로 인해 연관성을 찾는 데 어려움이 있었음.

추가 조사 및 분석 작업의 부담

- 예상과 다른 결과를 분석하기 위해 해당 시기에 발생한 국내 및 해외 사건을 조사해야 했으며, 이로 인해 시간이 추가적으로 소요됨.
- 사건과 데이터 간의 연관성을 찾는 과정이 까다로웠음.

문제 해결을 위한 시각화 및 데이터 연결

- 데이터의 흐름과 변화를 이해하기 위해 그룹별 시각화를 활용하고, 시기별 주요 사건과 유학생 수 변화를 연결하는 방식으로 문제를 해결.
- 이 과정에서 시간이 소요되었으나, 유학생 수 변화에 영향을 주는 원인을 파악하는 데 성공.

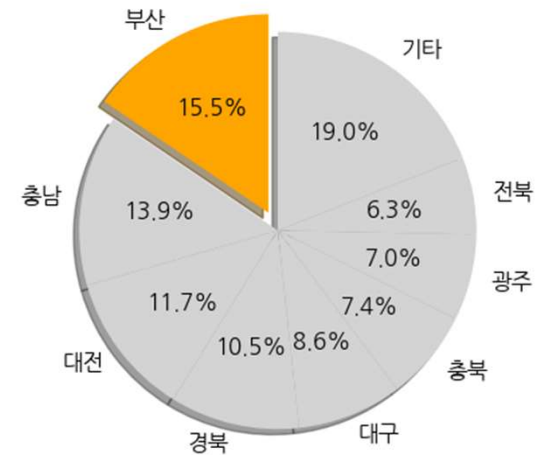
주제 선정 과정의 아쉬움

- 팀원들이 주제 선정 과정에서 적극적으로 의견을 제시했지만, 서로의 주제에 대해 의문을 제기하거나 추가 가설을 논의하는 데 부족함이 있었음.
- 이로 인해 결과물을 완성하는 데 시간이 오래 걸렸던 점이 아쉬움.

초기 가설 설정과 분업화의 한계

- 프로젝트 초기에는 각자 맡은 가설을 중심으로 작업을 진행했으나, 가설을 정리하고 핵심을 먼저 완성하는 데 집중하지 못했던 점이 아쉬움.
- 가설을 축소하고 역할을 분업화하여 효율적으로 작업했다면, 더 명확하고 빠르게 결과물을 완성할 수 있었을 것으로 판단.

2014년도 시도별 전체 외국인 유학생 비율 (수도권 제외, 5% 미만 기타)



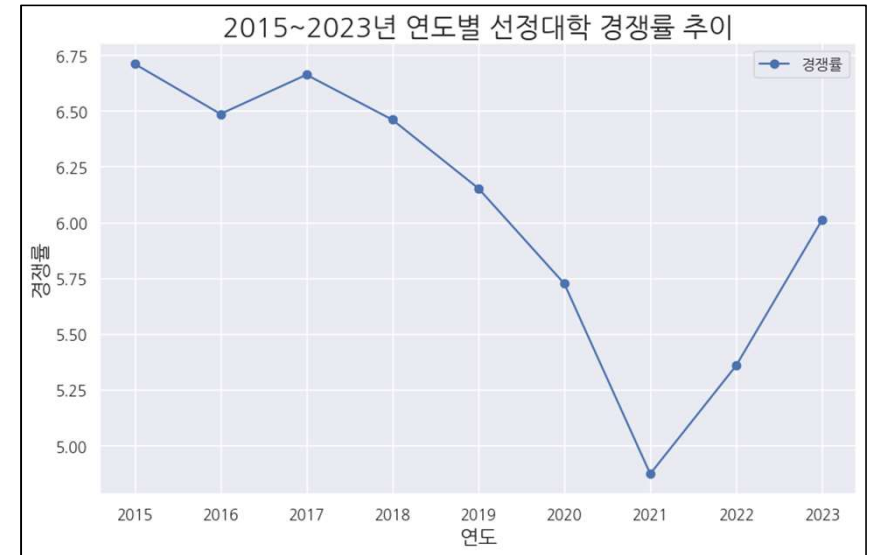
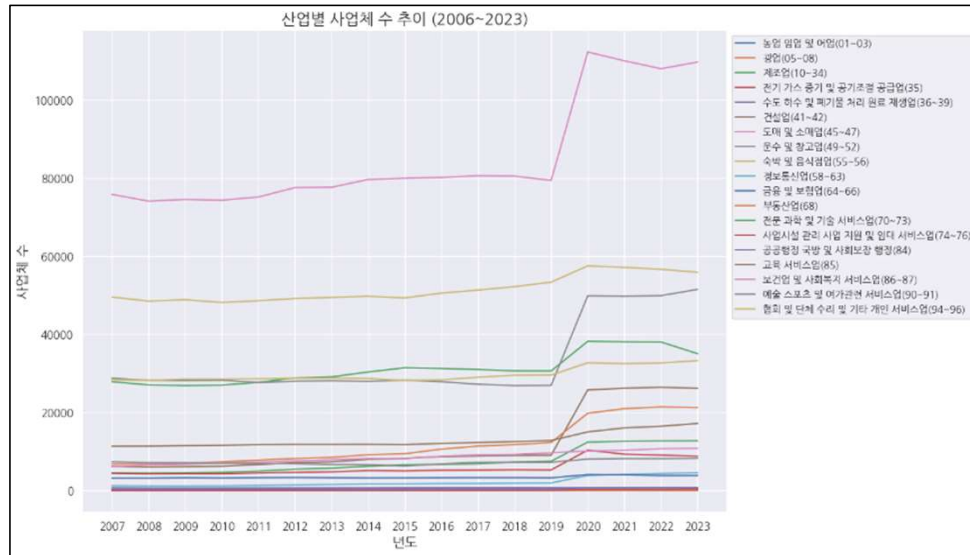
```
def animate(frame):
    ax.pie()

from matplotlib.animation import FuncAnimation
fig, ax = plt.subplots()
ani = FuncAnimation(fig, animate, frames=len(std))

# GIF로 저장
ani.save('./pie_plot.gif', fps= 0.5)
```

Part 4

한계 및 발전사항 : 이혁재



- 실제 경제 활동 가능 인구가 갈 수 있는 일자리 수의 구체화를 시키기 어려워 총 산업체 수를 확인하였으나 전체적으로 증가하는 추이로 확인되어 가설 근거로 사용하기 어려움이 있었음

- 인구분포와 함께 부산 사립대 경쟁률의 상관관계를 분석해보았으나 21년도 이후 상승에 대한 근거를 확인하지 못해 가설 근거로 사용하기 어려움이 있었음

Q & A

감사합니다.