

title.ipynb

```
from donggukUniversity import dataScience  
print(dataScience.course)
```

**데이터시각화**

```
print(topic)
```

**다양한 시각으로 바라본 전세계 여성의 삶**  
(폭력, 교육, 자살, 출산, 사회 진출, 경제를 중심으로)

```
df = data.about.us()  
df.iloc[0, :]
```

**정보통신공학과 2018112168 권나영**

```
df.iloc[1, :]
```

**통계학과 2019113421 최은진**

## | Index

- Intro
- Issue1 : 여성이 남편의 폭력이 정당하다고 생각하는 것은 교육 수준과 관련 있을까?
- Issue2 : 폭력의 정당화와 여성의 자살 수가 관련 있을까?
- Issue3 : 청소년의 출산 수와 총 여성의 출산 수는 상관 관계가 있을까?
- Issue4 : 여아의 취학률과 청소년의 출산 수는 상관 관계가 있을까?
- Issue5 : 여성의 피임률과 사회 진출은 상관 관계가 있을까?
- Issue6 : 여성의 사회 진출과 국가의 경제 발전은 상관관계가 있을까?
- Conclusion

## | Intro

- 올해 중순 아프가니스탄 내전으로 특별 기여자 체류 허용으로 입국 했다는 뉴스를 보고 관심을 가진 국가.
- 아프가니스탄의 여성들은 남성 동반자 없이 외출 금지, 교육과 경제 활동도 금지.
- 여성이 최초로 선거 참가권을 얻은 지 약 120년, 아직도 전세계적으로 보면 평등한 사회는 먼 것 같음.
- 다양한 시각으로 전세계 여성의 삶에 대해 분석해보고 싶었음.



얼굴 드러난 사진에 페인트칠 아프가니스탄 수도 카불 거리에서 한 남성이 웨딩 광고 속에 나온 여성 사진을 흰 페인트로 덧칠하고 있다. 톨로뉴스TV 대표 트위터 캡처

출처 : 경향신문 - 카불거리, 여성들이 사라졌다 (21.8.17)

## | Issue1

여성이 남편의 폭력이 정당하다고 생각하는 것은 교육 수준과 관련 있을까?

- 사용한 지표

지표 코드	설명
SG.VAW.ARGU.ZS	'남편과 논쟁 시 맞아도 된다'고 믿는 여자의 비율
SG.VAW.BURN.ZS	'요리를 태웠을 때 남편에게 맞아도 된다'고 믿는 여자의 비율
SG.VAW.GOES.ZS	'말 없이 외출 시 남편에게 맞아도 된다.'고 믿는 여자의 비율
SG.VAW.NEGL.ZS	'아이들을 무시하면 남편에게 맞아도 된다'고 믿는 여자의 비율
SG.VAW.REFU.ZS	'성관계를 거부하면 남편에게 맞아도 된다'고 믿는 여자의 비율
SE.TER.CUAT.BA.MA.ZS	25세 이상 남성 중 '학사' 학위 이상을 가진 남자의 비율
SE.TER.CUAT.BA.FE.ZS	25세 이상 여성 중 '학사' 학위 이상을 가진 여자의 비율

- 데이터 전처리 과정

각 지표에 관한 데이터를 라이브러리를 이용해 전체 값이 없는 행은 제외하고 DataFrame 형식으로 저장.

5가지 이유의 폭력의 정당화에 관한 데이터는 결측치가 이미 많았음

▶ 제거하거나 대체하는 대신, 존재하는 데이터에 대해서만 대륙, 국가, 년도, 값을 정리 후 분석.

국가별 남녀 학사 이상 비율 데이터는 폭력의 데이터가 있는 나라에 대해서만 가지고 온 후 두가지로 전처리.

- 1) 전체 기간에 대한 대륙별 평균

- 2) 폭력의 정당화 데이터의 국가, 년도와 대응하는 데이터만 가지고 와서 데이터프레임을 연결 후 분석

## | Issue1

여성이 남편의 폭력이 정당하다고 생각하는 것은 교육 수준과 관련 있을까?

- '대륙' 데이터?

wbgapi 에서 economy 데이터들을 데이터프레임으로 불러오기

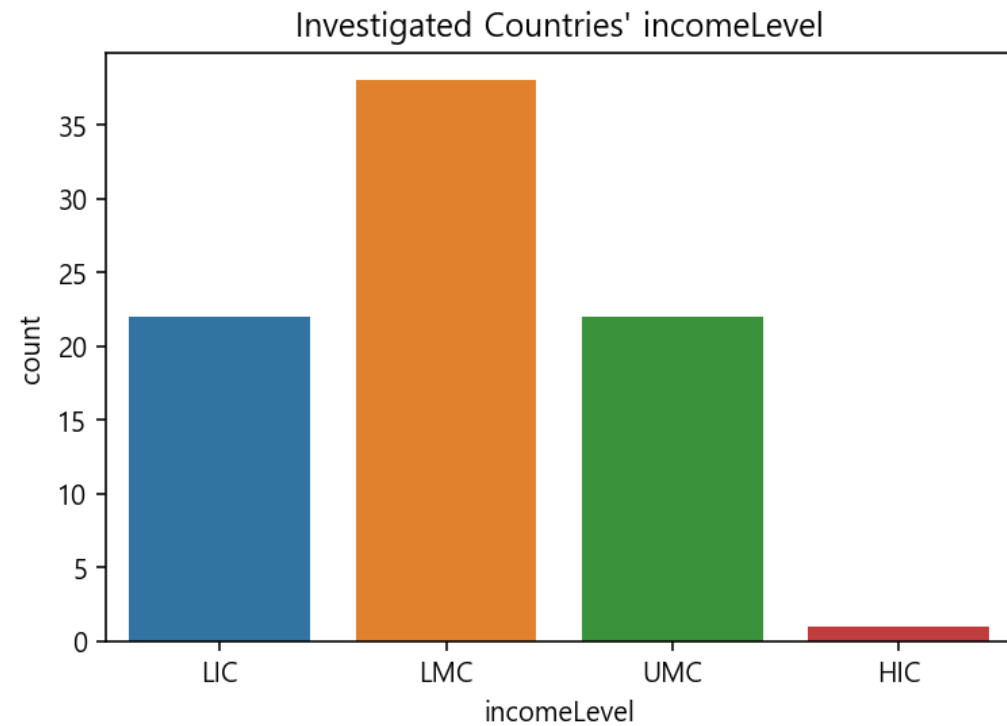
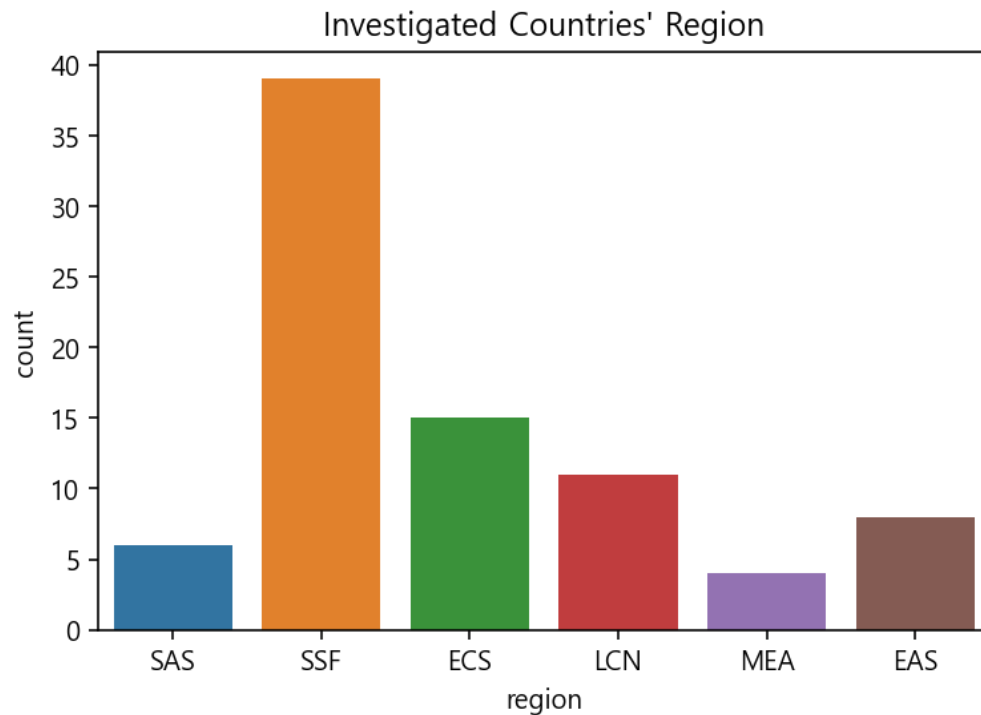
▶ wbgapi.economy.dataframe.reset\_index()

	id	name	aggregate	longitude	latitude	region	adminregion	lendingType	incomeLevel	capitalCity
0	ABW	Aruba	False	-70.0167	12.51670	LCN		LNK	HIC	Oranjestad
1	AFE	Africa Eastern and Southern	True	NaN	NaN					
2	AFG	Afghanistan	False	69.1761	34.52280	SAS	SAS	IDX	LIC	Kabul
3	AFW	Africa Western and Central	True	NaN	NaN					
4	AGO	Angola	False	13.2420	-8.81155	SSF	SSA	IBD	LMC	Luanda
...	...	...	...	...	...	...	...	...	...	...
261	XKX	Kosovo	False	20.9260	42.56500	ECS	ECA	IDX	UMC	Pristina
262	YEM	Yemen, Rep.	False	44.2075	15.35200	MEA	MNA	IDX	LIC	Sana'a
263	7AF	South Africa	False	28.1871	-25.74600	SSF	SSA	IBD	UMC	Pretoria

## | Issue1

여성이 남편의 폭력이 정당하다고 생각하는 것은 교육 수준과 관련 있을까?

- EDA 1차 : 각 국가별 존재하는 데이터는 적기 때문에 countplot 함수를 이용해 대륙별, 소득 수준별로 존재하는 데이터의 개수에 대하여 분석

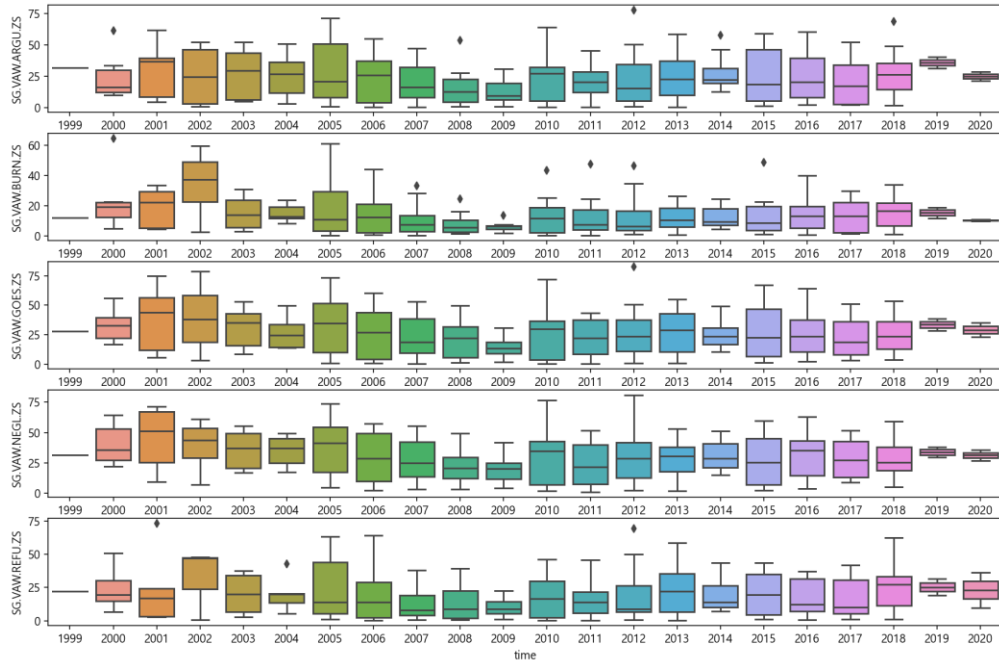


- 결과 : 사하라 이남의 아프리카 대륙과 중하위소득의 데이터가 가장 많음.

## | Issue1

여성이 남편의 폭력이 정당하다고 생각하는 것은 교육 수준과 관련 있을까?

- EDA 2차 : 남편의 폭력을 정당하다고 생각하는 여성 비율의 연도별 추이



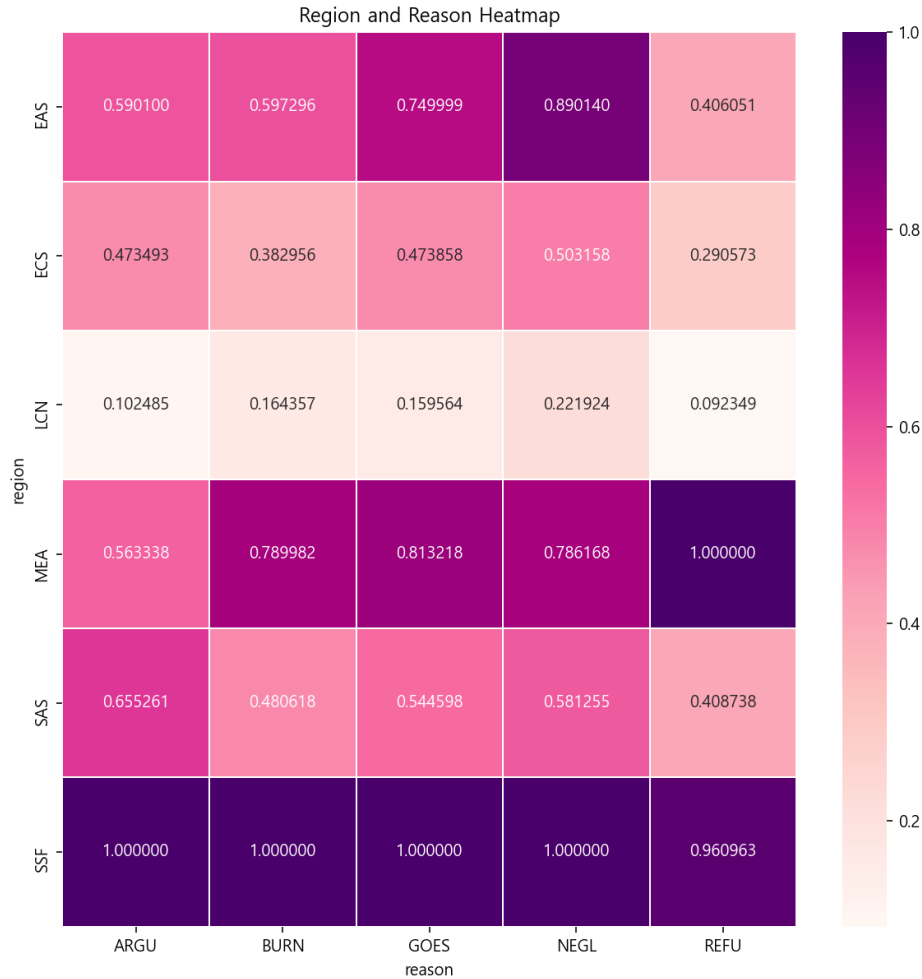
- 결과 : 5가지 폭력 이유에 대해서 최대값이 대부분 00년대에 나타났으며, 이상치로 인해 다양한 값이 나와서 연도별 모든 국가의 평균값으로 다시 분석. 2002년쯤이 가장 높았지만 현대에 와서 눈에 띄게 줄어들지는 않았음. 아이의 무시에 대한 폭력이 정당하다고 생각하는 경향.



## | Issue1

여성이 남편의 폭력이 정당하다고 생각하는 것은 교육 수준과 관련 있을까?

- EDA 3차 -1 : 대륙별 폭력 이유에 대한 평균을 최대값으로 나눈 후 상관 관계 분석



- 결과

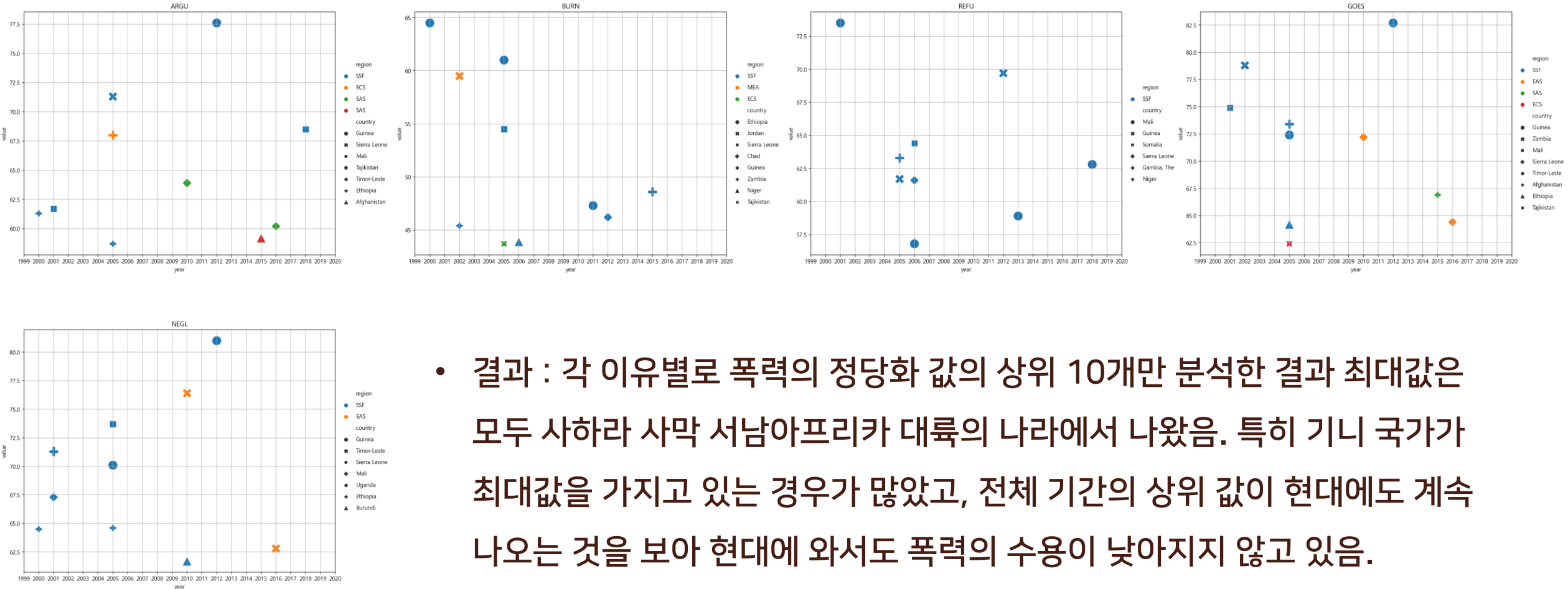
중남미 대륙이 어느 이유든 폭력의 정당화 된다고 생각하지 않으며, 사하라 사막 이남의 아프리카 지역은 폭력이 정당하다고 생각하는 편이다.

중동과 북아프리카에서 성관계 거부 시 폭력에 대해서만 다른 대륙과 비교하여 높은 편이다.

# | Issue1

여성이 남편의 폭력이 정당하다고 생각하는 것은 교육 수준과 관련 있을까?

## • EDA 3차 -2 : 이유별로 상위 10개 값의 국가와 값 분포

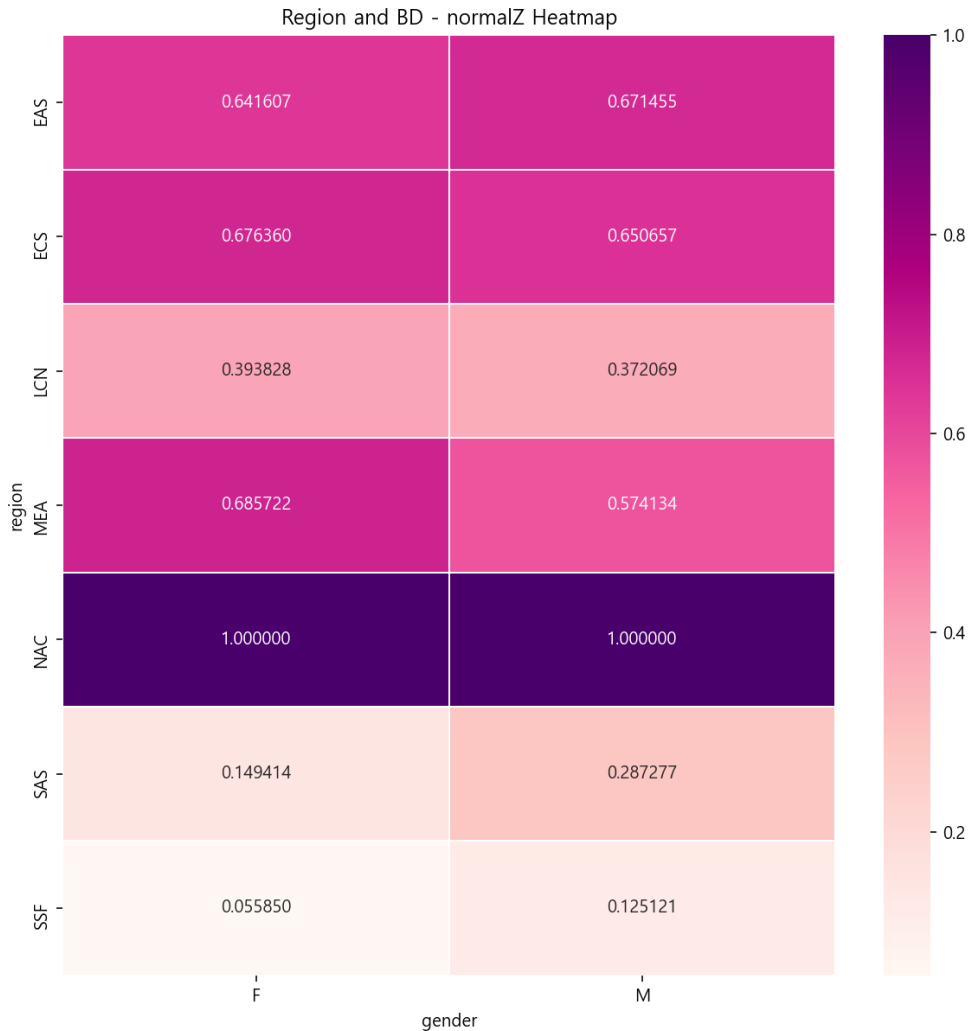


- 결과 : 각 이유별로 폭력의 정당화 값의 상위 10개만 분석한 결과 최대값은 모두 사하라 사막 서남아프리카 대륙의 나라에서 나왔음. 특히 기니 국가가 최대값을 가지고 있는 경우가 많았고, 전체 기간의 상위 값이 현대에도 계속 나오는 것을 보아 현대에 와서도 폭력의 수용이 낮아지지 않고 있음.

## | Issue1

여성이 남편의 폭력이 정당하다고 생각하는 것은 교육 수준과 관련 있을까?

- EDA 4차 : 대륙별 남녀 학사 이상의 비율의 양상을 히트맵으로



- 결과

남녀 학사 비율에 대한 데이터를 대륙별 평균값으로 본 결과 북미 지역의 남녀 학사 취득률이 가장 높고 사하라 서남 지역의 학사 취득률이 낮으며, 이 대륙의 여성의 학사 취득 비율은 북미 지역을 남, 여 각각 1로 두었음에도 남성 비율의 절반 수준임.

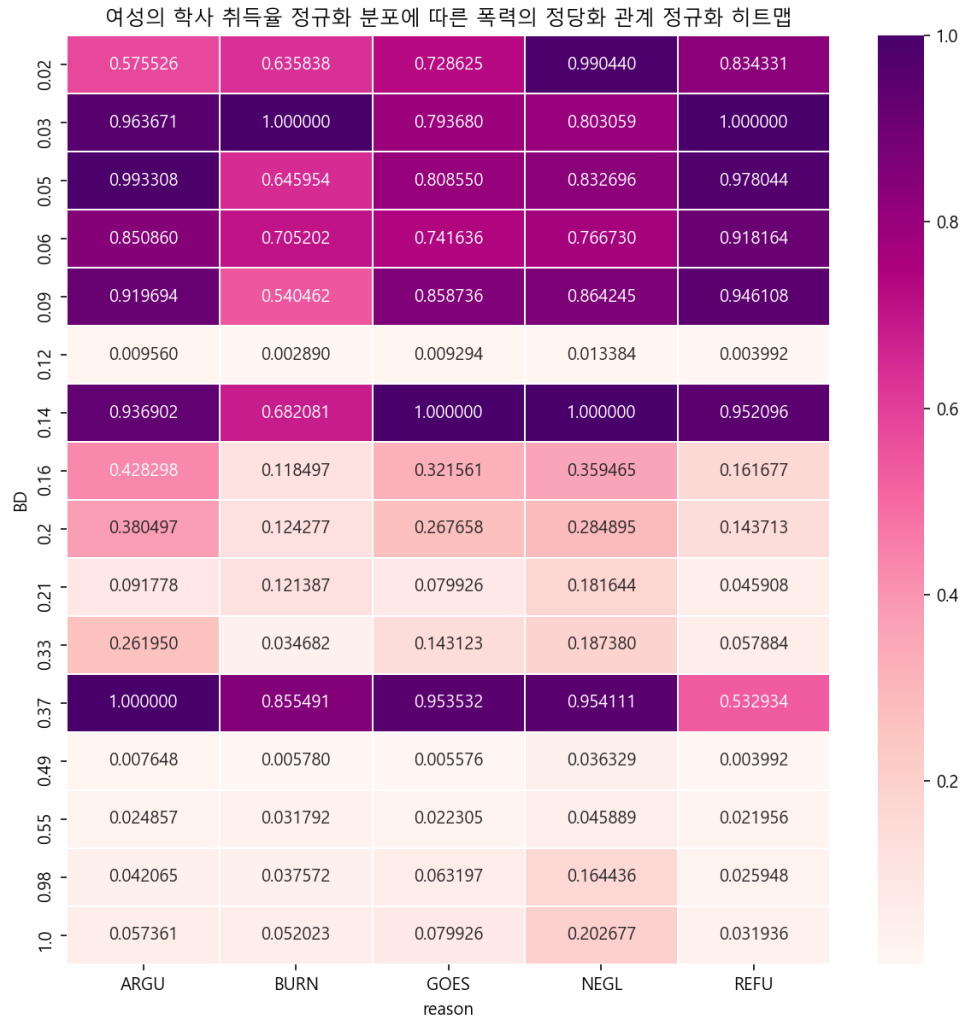
- 결론

남편의 폭력을 정당화 하는 대륙의 여성,남성 학사 취득률이 가장 낮음.

# | Issue1

여성이 남편의 폭력이 정당하다고 생각하는 것은 교육 수준과 관련 있을까?

- EDA 5차 : 여성의 학사 취득율과 이유별 폭력의 정당화에 대한 관계를 히트맵으로



- 결과 : 폭력에 관한 데이터가 존재하는 국가, 연도의 여성의 학사 취득률을 추출 후 히트맵 분석. 학사 취득률의 최대값을 1로 두어 정규화 한 후, 폭력 이유별 폭력의 정당화 값을 바라본 결과 학사 취득률이 낮을 수록 남편의 폭력을 수용하는 경향
- 결론 : 데이터가 적어서인지 정확하진 않지만 예상처럼 여성의 학사 이상 취득률과 폭력의 정당화는 관계가 있고 반대 관계임.

## • 사용한 지표

지표 코드	설명
SG.VAW.ARGU.ZS	'남편과 논쟁 시 맞아도 된다'고 믿는 여자의 비율
SG.VAW.BURN.ZS	'요리를 태웠을 때 남편에게 맞아도 된다'고 믿는 여자의 비율
SG.VAW.GOES.ZS	'말 없이 외출 시 남편에게 맞아도 된다.'고 믿는 여자의 비율
SG.VAW.NEGL.ZS	'아이들을 무시하면 남편에게 맞아도 된다'고 믿는 여자의 비율
SG.VAW.REFU.ZS	'성관계를 거부하면 남편에게 맞아도 된다'고 믿는 여자의 비율
SH.STA.SUIC.FE.P5	여성 100,000명당 자살수

- 데이터 전처리 과정

Issue1와 동일한 방법으로 전체 값이 없는 행은 제외하고 DataFrame 형식으로 저장.

10만명당 여성의 자살수 데이터는 폭력 데이터가 있는 국가의 데이터만 가지고 와서 두가지 방법으로 전처리

1) 대륙별 자살수 평균 값 계산

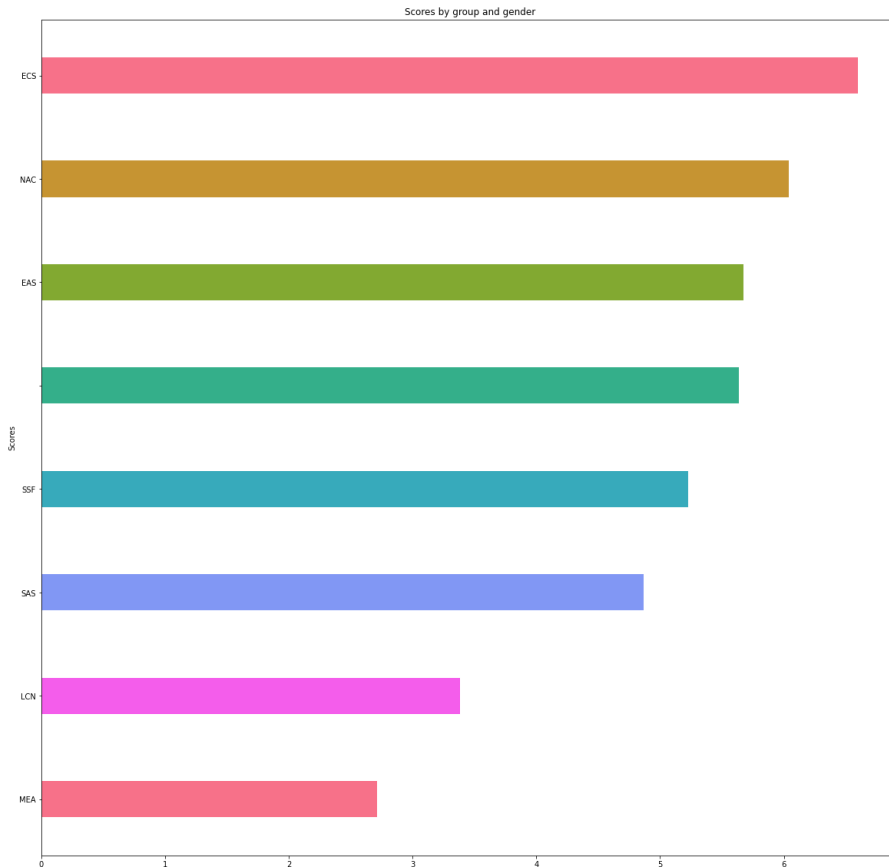
2) 5가지 이유의 남편의 폭력 정당화에 관한 데이터가 존재하는 년도의 데이터만 남기고

그 자살 수를 10단계로 나누어 상관관계 분석

## | Issue2

### 폭력의 정당화와 여성이 자살 수가 관련 있을까?

- EDA 1차 : 여성 10만명당 자살 수의 대륙별 평균을 subplot으로



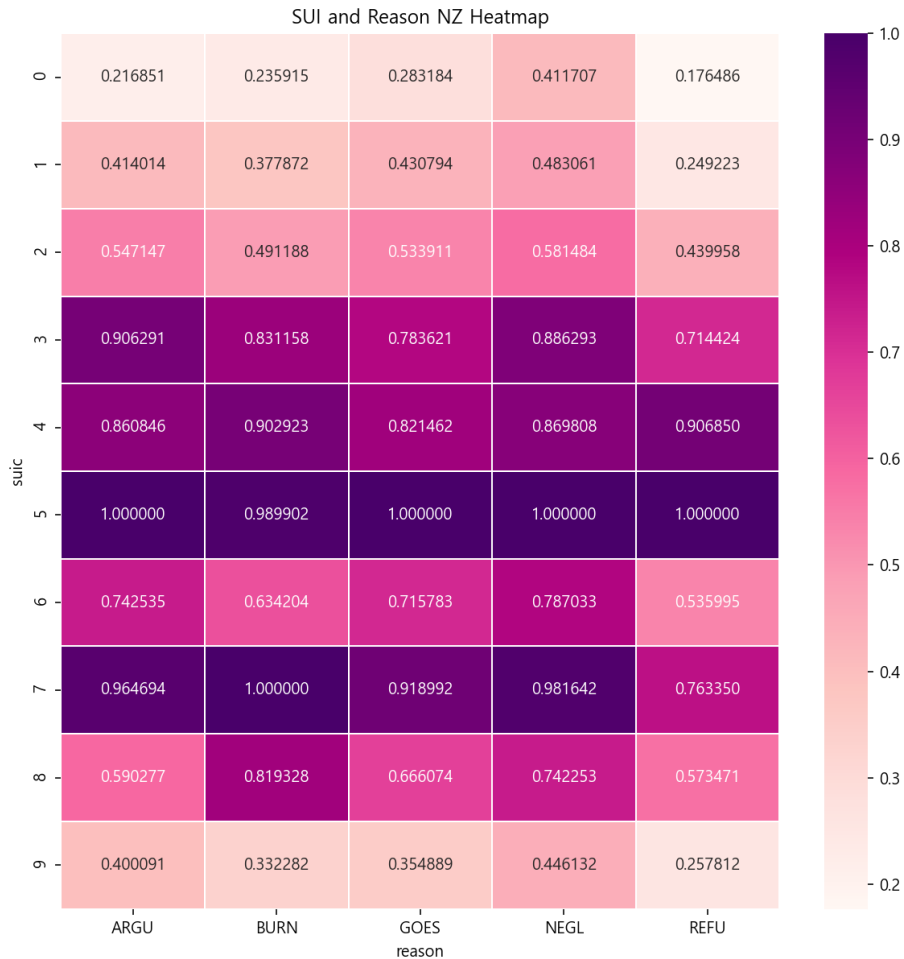
- 결과

자살 수 평균이 유럽과 중앙 아시아에서 가장 높고  
중동과 북아프리카가 가장 낮음.

- 결론

유럽과 중앙아시아의 자살 수가 중동과 북아프리카의  
자살 수는 두배임.

- EDA 2차 : 여성 폭력의 정당화 비율과 여성 자살수 비교를 히트맵으로



### • 결과

데이터가 존재하는 값끼리 비교한 결과 폭력의 수용과 자살수의 유의미한 상관관계가 없음

### • 결론

남편의 폭력에 대한 정당화가 높을수록, 자존감 부재와 문화·도덕적 지식 습득의 결여를 동시에 겪을 것이라 예상하며 이러한 환경이 여성의 자살 수를 높게 할 것이라 생각했지만 오히려 유럽, 중앙아시아의 자살수가 높으므로 자살의 인과 관계는 다른 것이 있을 것이라 보임.



- 사용한 지표

지표 코드	설명
SP.ADO.TFRT	15-19세 여성의 1000명이 출산하는 아이의 수
SP.DYN.TFRT.IN	전체 나이의 여성 1명이 출산하는 아이의 수

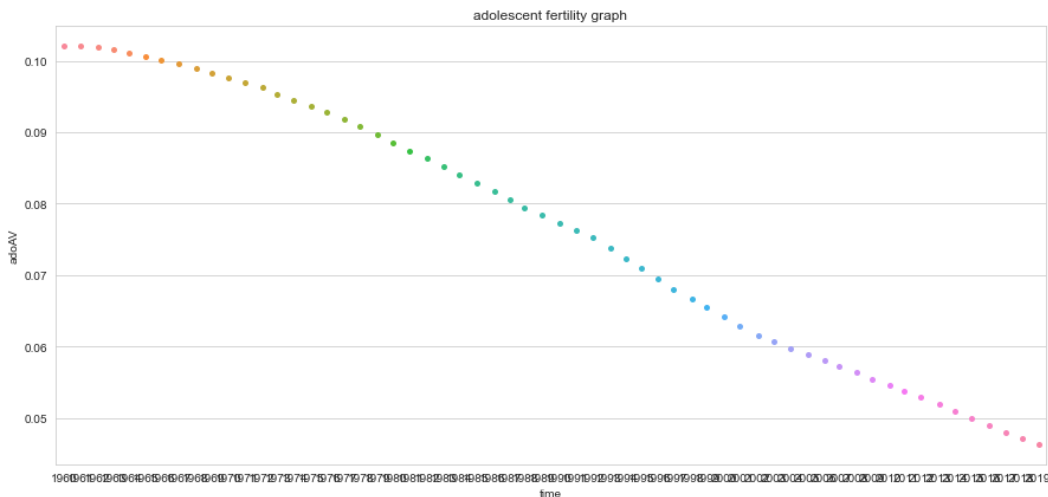
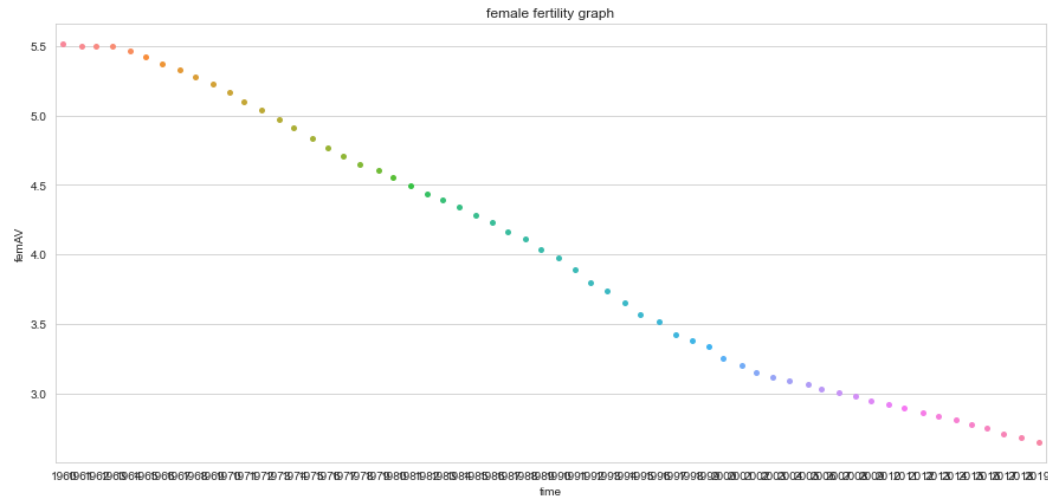
- 데이터 전처리 과정

전체 값이 없는 행은 제외하고 DataFrame 형식으로 각각 저장.

1) 두 데이터프레임 결합 후 컬럼 명을 처리하고 인덱스 초기화 후 전세계의 두가지 출산 수에 대해 연도별 평균 계산. 청소년은 1000명 당 출산 수이고 일반 여성은 1명당 출산 수이기 때문에 단위를 맞춰서 시계열 분석

2) 국가별로 청소년, 전체 여성 출산 수의 평균을 구하고, 그래프의 해석 편리를 위해 전체 여성 출산 수를 10단계로 범주화하여 여성 출산 수 10단계와 청소년의 출산 수 비교

- EDA 1차 : 전체 연도에 대해서 청소년과 여성 출산 수 비교를 strip plot으로



### 결과

연도에 반비례, 시간이 지날 수록 떨어지고 있다. 데이터가 수집된 초기 (60년대)와 비교하면 최근 출산 수는 약 절반 수준임.

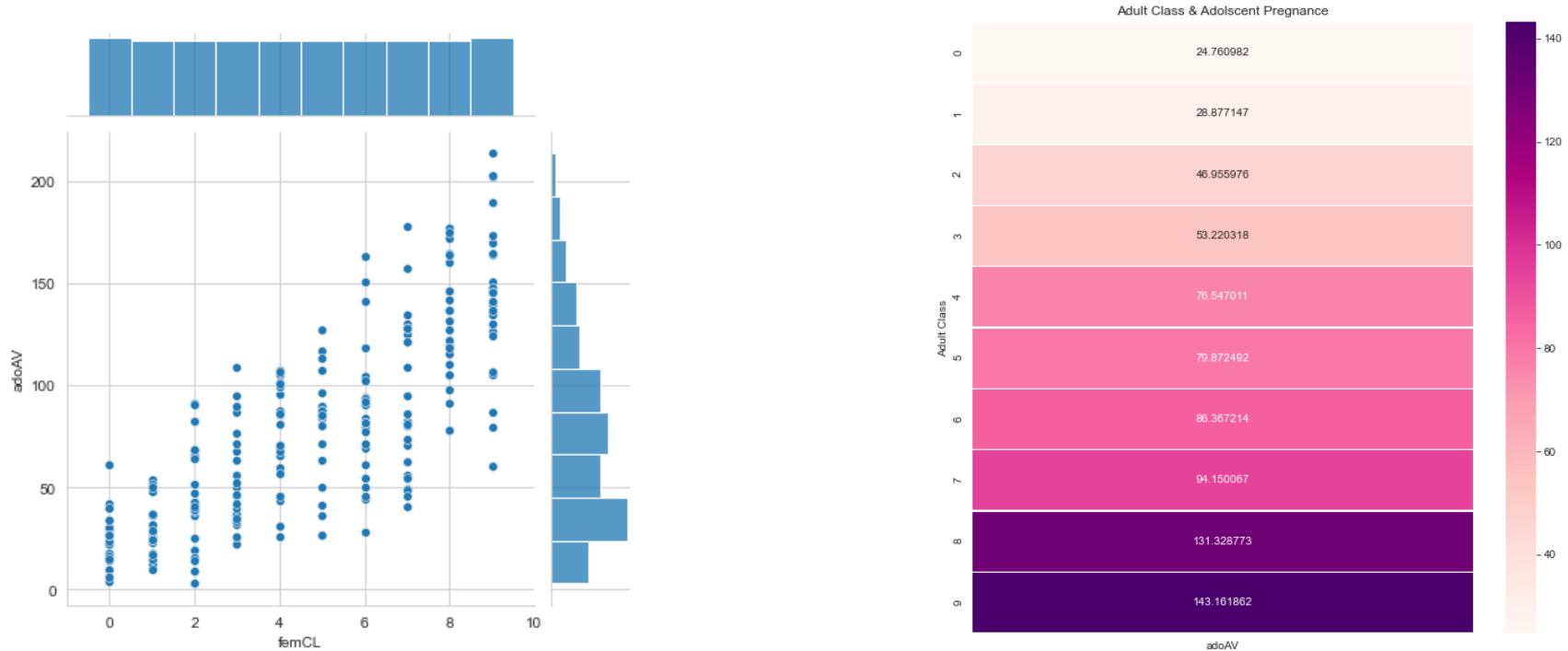
### 결론

전세계 인구는 폭발적으로 증가하고 있는데 출산 수는 감소하고 있는 것으로 보아 의학 발전으로 수명이 길어졌기 때문이라고 생각.

## | Issue3

청소년의 출산 수와 총 여성의 출산 수는 상관 관계가 있을까?

- EDA 2차 : 전체 연령의 여성 출산 수와 청소년 출산 수를 joint plot과 heat map으로 비교



- 결과 : 여성의 출산 수를 10단계로 나타내서 청소년 출산 수와 비교한 결과 양의 상관 관계이다.
- 결론 : 예상과 일치. 여성 1명의 출산 수가 높은 나라는 청소년 출산 수도 높은 편인 것을 확인

### • 사용한 지표

지표 코드	설명
SP.ADO.TFRT	15-19세 여성의 1000명이 출산하는 아이의 수
SE.ENR.PRSC.FM.ZS	초등 및 중등 교육에서 남아에 대한 여아의 비율(%)

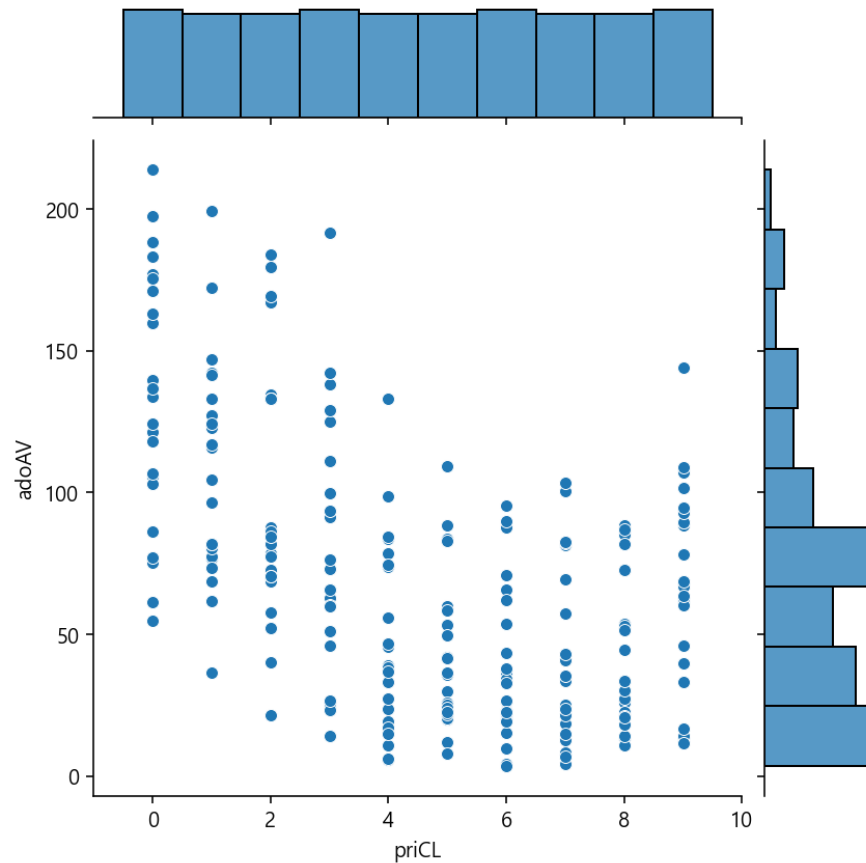
- 데이터 전처리 과정

전체 값이 없는 행은 제외하고 DataFrame 형식으로 저장.

나라별로 취학률과 청소년의 출산 수 평균을 계산

그래프 해석의 편리성을 위해 여아의 취학률은 10단계로 범주화 한 후, 청소년 출산 수와 비교

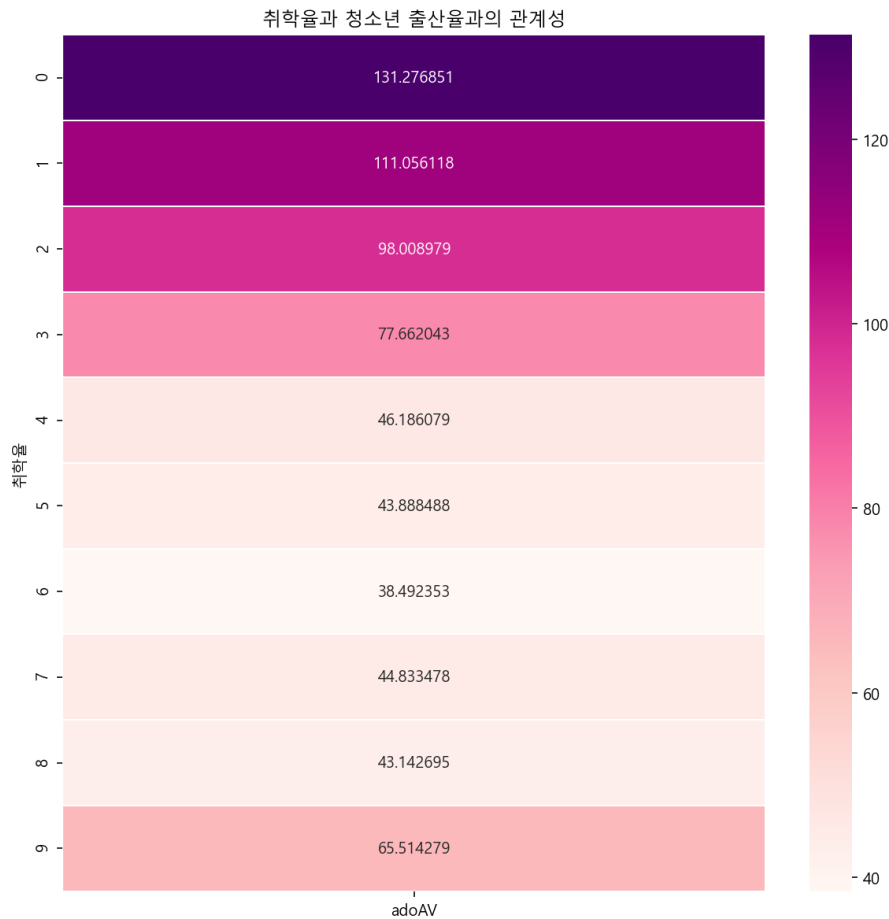
- EDA 1차 : 여아의 취학률과 청소년의 출산 수를 joint plot으로



### • 결과

대체적으로 넓게 분포되어 있지만 청소년 출산 수가 높으면 취학률이 낮은 편이고, 취학률이 높으면 청소년 출산 수는 낮은 편이다.

- EDA 2차 : 여아의 취학률과 청소년 출산 수를 heat map으로 비교



- 결과 : 취학률이 낮으면 출산 수가 높은 것은 확실. 하지만 취학률이 가장 높았을 때 청소년 출산 수가 낮지 않은 않은 의외의 경우도 있음.
- 결론 : 출산 때문에 학업을 연장하지 못하는 여아들이 많을 것이라 생각했는데 의외였음. 취학률은 초등학교-중학교 기준이고, 청소년 출산 수는 15-19세 기준이라 우리나라 기준으로 생각하면 고등학교 진학을 포함하지 못한 것으로 보임. 고등학교 포함 시 다른 결과가 나왔을 수도? 아니면 복지의 효과?



- 사용한 지표

지표 코드	설명
SP.DYN.CONU.ZS	15-49세 여성의 모든 방법의 피임 보급률(%)
SG.GEN.PARL.ZS	전체 국회의원 중 여성 의석 비율(%)
SL.AGR.EMPL.FE.ZS	모든 여성 노동자 중 농업계 종사율(%)
SL.IND.EMPL.FE.ZS	모든 여성 노동자 중 산업계 종사율(%)
SL.SRV.EMPL.FE.ZS	모든 여성 노동자 중 서비스업계 종사율(%)

- 데이터 전처리 과정

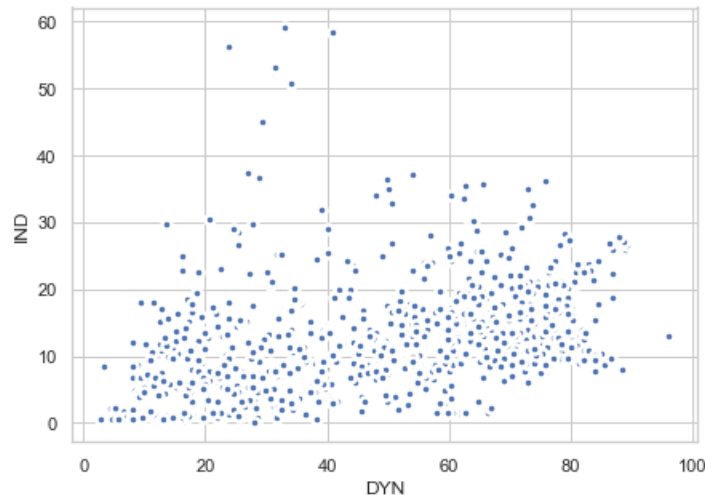
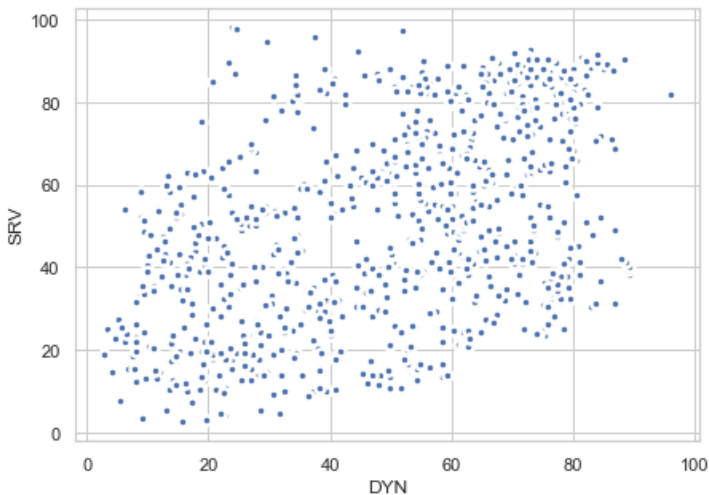
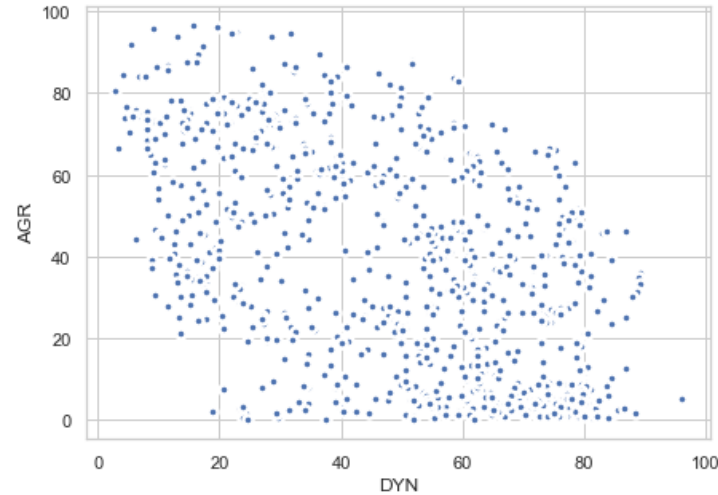
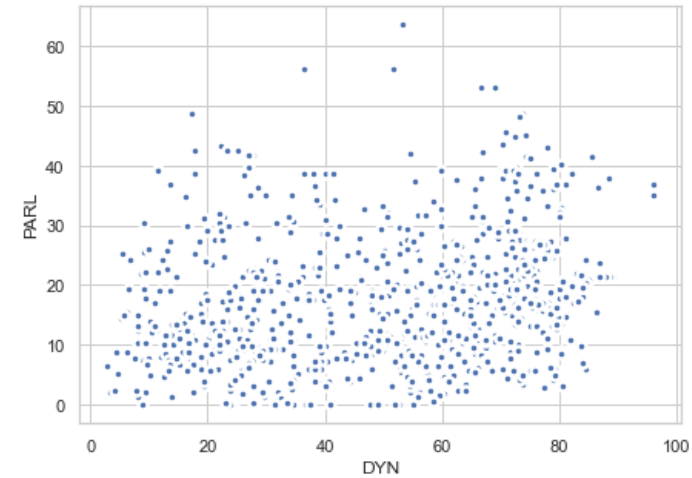
전체 값이 없는 행은 제외하고 DataFrame 형식으로 저장

결측 값이 많아서 5가지 지표에 대한 값이 모두 채워진 국가, 연도 행만 남기고 삭제 후 행 번호 초기화

1) 위의 데이터 그대로 사용

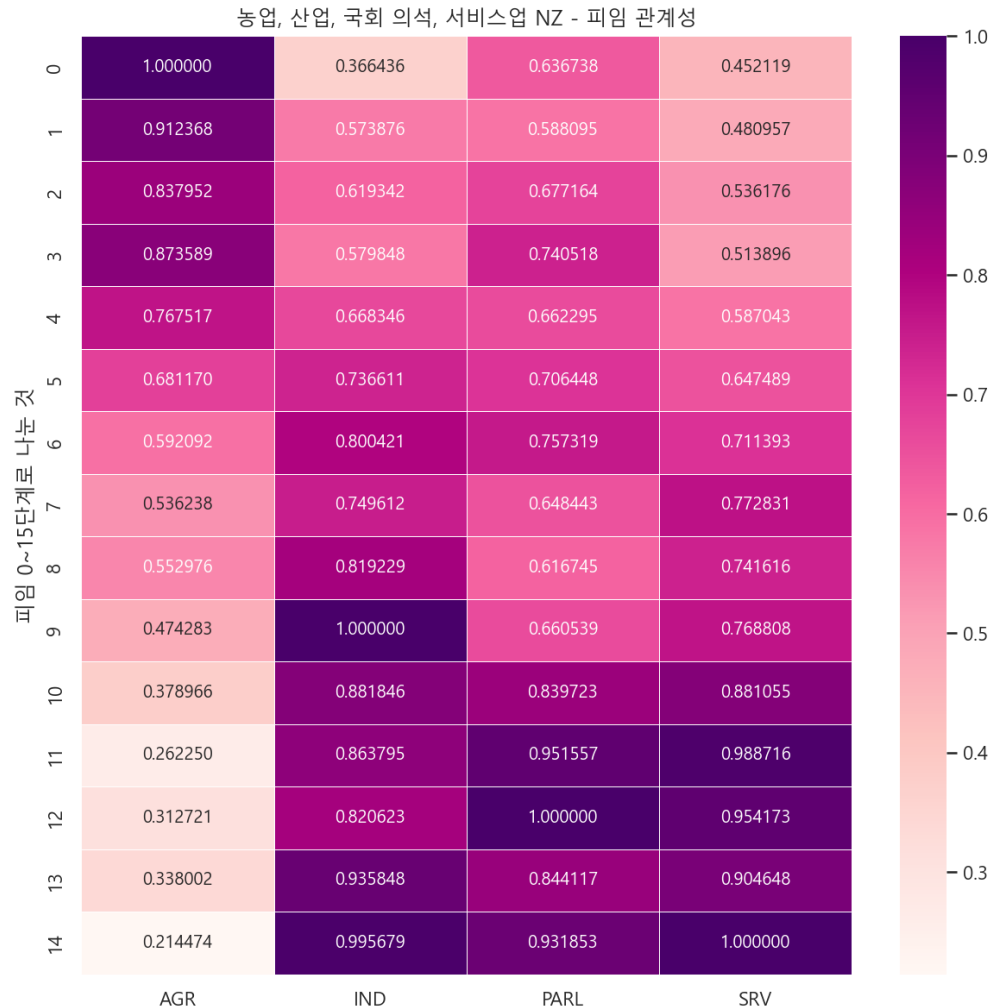
2) 그래프 해석 편의를 위해 피임률을 0~15단계로 범주화 하여 활용

- EDA 1차 : 여성 전체 노동자 중 해당 분야 종사율과 피임 보급률 scatter plot으로 비교



- 결과 : 모두 제각각 다른 분포 모양을 보임. 농업만 눈에 띄게 피임 보급률이 낮을 수록 종사율이 높음.
- 결론 : 농업에 종사하면 육체 노동 강도가 높아서 임신할 경우 일을 하기가 쉽지 않을텐데 피임률이 낮은 이유가 궁금함.

- EDA 2차 : 여성 전체 노동자 중 해당 분야 종사율과 피임 보급률 heat map으로 비교



- 결과 : 여성 노동자가 서비스직에 많이 종사하는 국가일수록 피임이 보편화 되어있고 농업에 많이 종사하는 국가는 피임률이 낮다.
- 결론 : 여성이 국회 의원을 차지하는 비율은 전체 의원 중 여자가 차지하는 비율인데, 나머지 고용률은 전체 여성 노동자 수 중 해당 산업 종사율을 모델링한 추정치이기 때문에 해당 분야의 전체 노동자 중에서 여성이 차지하는 비율이었으면 더 비교하기 쉬웠을 것 같음.

## • 사용한 지표

지표 코드	설명
SE.ENR.PRSC.FM.ZS	초등 및 중등 교육에서 남아에 대한 여아의 비율(%)
SP.DYN.CONU.ZS	15-49세 여성의 모든 방법의 피임 보급률(%)
SG.GEN.PARL.ZS	전체 국회의원 중 여성 의석 비율(%)
SL.AGR.EMPL.FE.ZS	모든 여성 노동자 중 농업계 종사율(%)
SL.IND.EMPL.FE.ZS	모든 여성 노동자 중 산업계 종사율(%)
SL.SRV.EMPL.FE.ZS	모든 여성 노동자 중 서비스업계 종사율(%)
NY.GDP.PCAP.CD	1인당 GDP(US\$)

- 데이터 전처리 과정

전체 값이 없는 행은 제외하고 DataFrame 형식으로 저장

결측 값이 많아서 7가지 지표에 대한 값이 모두 채워진 국가, 연도 행만 남기고 삭제 후 분석

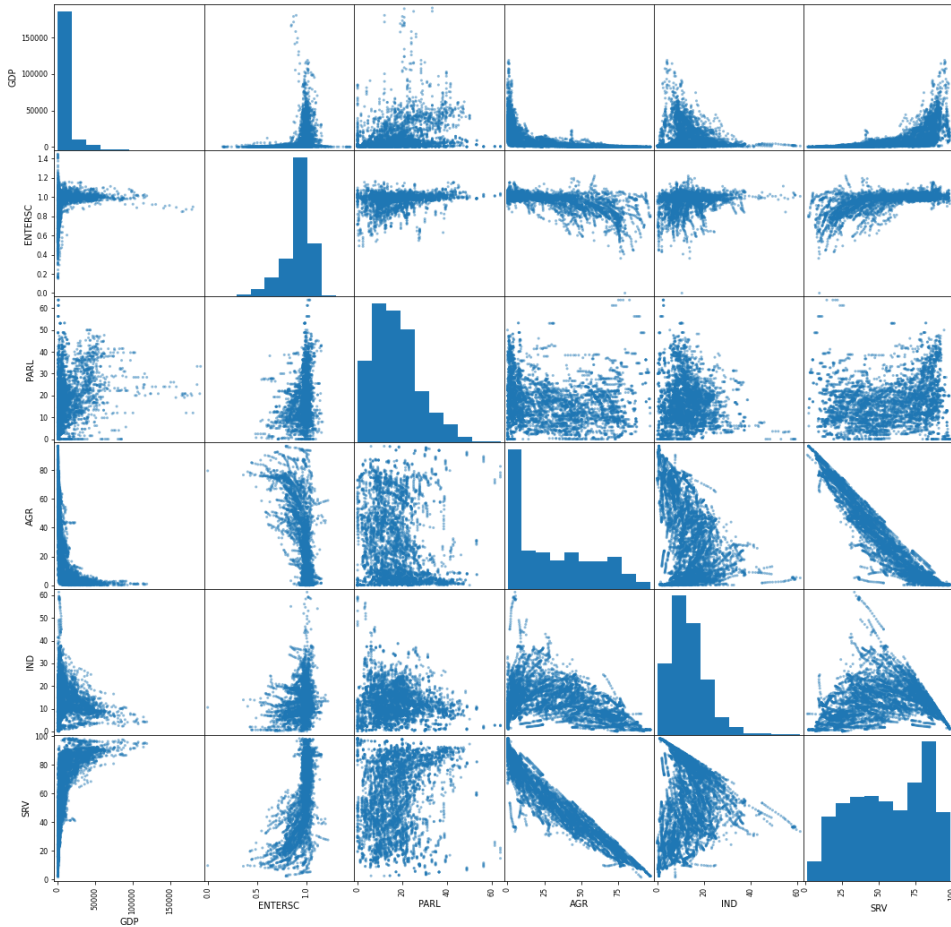
1) 위의 데이터 그대로 사용

2) 년도별로 각 지표에 대해 평균을 계산하고 각 지표들의 범위가 다르기 때문에

각 지표의 최대값을 1로 스케일링

3) 1인당 GDP를 15단계로 범주화 하여 분석

- EDA 1차 : 각 지표 간 관계를 보기 위해 scatter matrix로 시각화

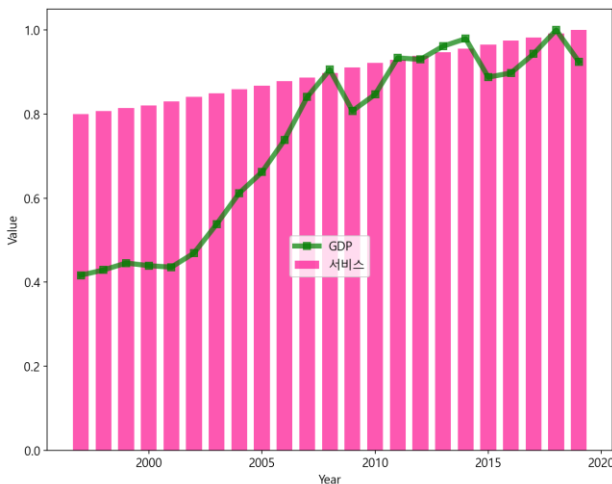
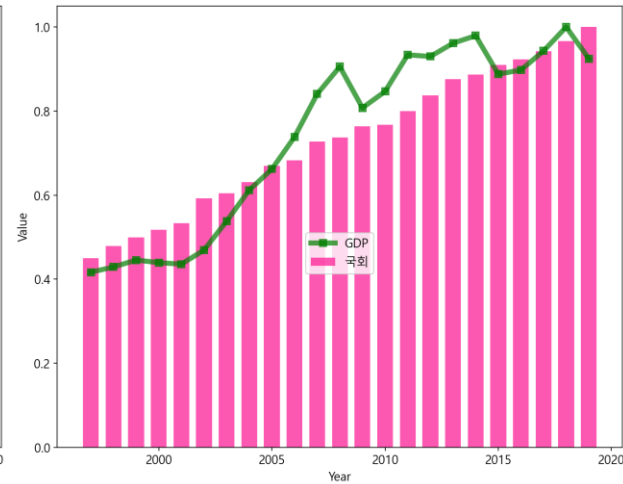
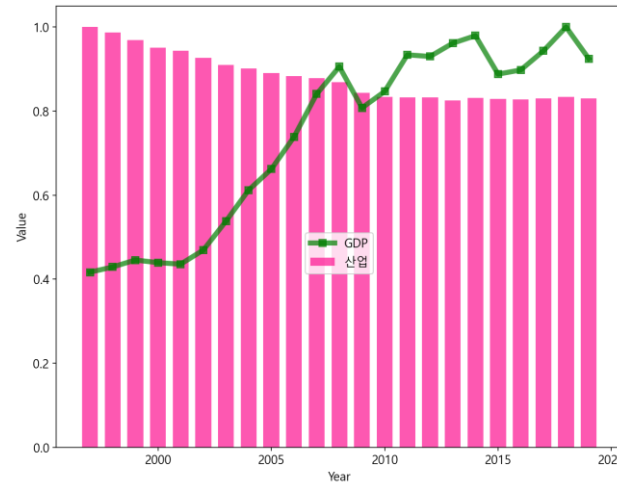
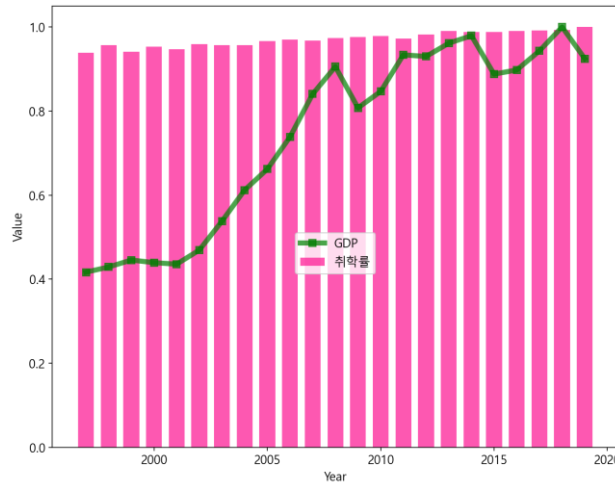
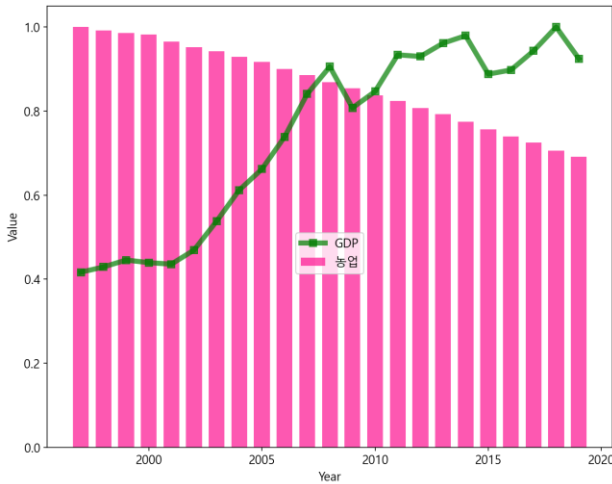


- 결과

농업 종사율과 서비스직 종사율은 뚜렷한 음의 상관관계임.

각 지표들은 정규분포와 거리가 멀어 보임.

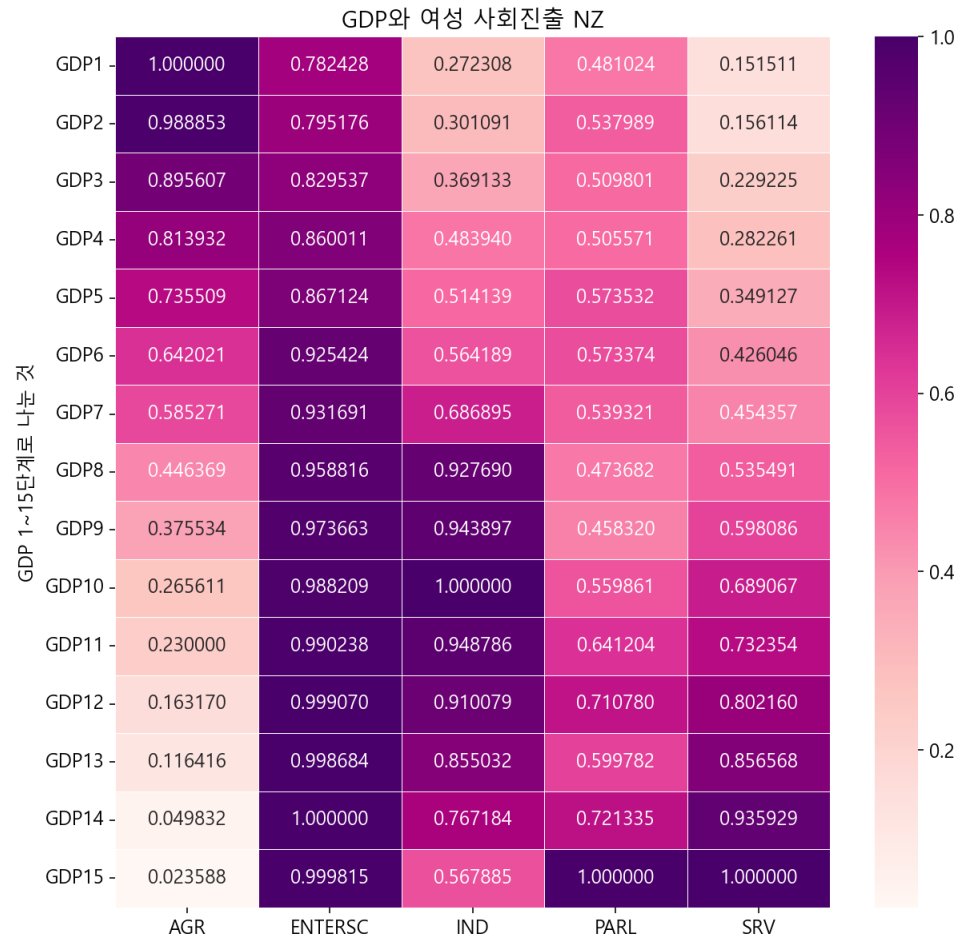
- EDA 2차 : 교육, 각 분야의 종사율과 1인당 GDP를 막대, 꺾은선 그래프 시계열 분석



- 결과 : 시간의 흐름에 따라 각 지표 값의 변화를 분석한 결과  
1인당 GDP는 우상향 추세. 여성의 농업 종사율은 떨어지고 있음. 최대 종사율에 대비하여 약 70% 수준. 취학률은 미세하지만 꾸준히 높아지고 있음. 산업 종사율이 떨어지고 있는게 의외. 여성의 국회 의석 비율은 급격히 상승 중임. 20년간 약 2배 상승.



- EDA 3차 : 1인당 GDP를 15단계로 나누어 각 지표와 관계를 heat map으로 비교



### • 결과

여성이 농업 분야에서 많이 일하는 나라는 아직 경제적으로 발전해야 할 국가. GDP 10단계에서 산업 종사율이 최대이며 급격히 감소. 진학률, 서비스업 종사율, 국회 의원 비율은 1인당 GDP와 정확히 비례함!

### • 결론

예상처럼 국가가 경제적으로 발전한 나라는 여아가 교육을 많이 받고, 여성 국회의원 비율도 높다.

## | Conclusion

다양한 시각으로 전세계 여성의 삶을 분석한 EDA를 마치며...

- 느낀 점

편견이나 막연한 생각들을 객관적인 데이터로 직접 시각화 한 후 검증할 수 있는 시간이었습니다. 결측치가 많아 아쉬운 점이 많았지만 존재하는 데이터로만 분석하는 것도 충분히 의미 있었습니다. 다음 업데이트가 있을 때는, 남편의 폭력을 수용하는 여성의 수가 낮아질 수 있기를 바랍니다.

- 다음에 분석해 보고 싶은 주제

- 1) GDP의 증가와 인플레이션의 관계

- 2) 자살수가 높은 국가를 선정하여 자살률과 양의 상관 관계를 가지는 지표는 무엇인지, 그리고 인과관계가 있는지 궁금합니다.

- 3) 전세계 여성의 삶에 대한 지표들을 바라보았으니 다음에는 한 국가, 특히 한국에 대해서 깊이 분석해보고 싶습니다. 우리나라의 경우 남아의 수와 비교하여 여아의 진학률이 높은 편인데 국회의원 수는 전세계 평균보다 적어서 시간이 지나면 높아질지 궁금합니다.

bye.ipynb

```
if __name__ == "__end__":
```

```
    say()
```

```
def say() :
```

```
    if page is last:
```

```
        print("Thank you")
```