

데이터분석및시각화 과제2 발표

GDP는 국가의 행복을 반영하는가?

정오바기 팀

정지훈 오광혁 박영웅 이윤주

CONTENTS

- 1** 서론
- GDP란?
 - GDP가 높은 국가일수록 행복할까?

- 2** 본론
- 사용 데이터
 - 데이터 전처리
 - 데이터 시각화

- 3** 결론
- 결과 분석
 - 대안 제시
 - 최종 결과

GDP 란?

> GDP (Gross Domestic Product)

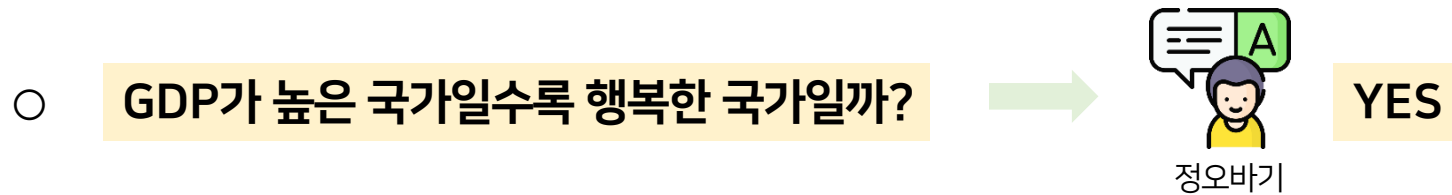
- 국내 총 생산
- 일정 기간에 한 국가 안에서 생성된 최종 생산물에 대한 시장 가치의 합
 - 2021년 대한민국 GDP: 2021년 한 해 동안 우리나라 국경 안에서 생산된 모든 재화와 서비스의 가격을 합한 금액
- GDP 항목: 교육 서비스, 운송 배달, 농장 수확, 공장 생산 ...
- GDP 기준



GDP는 한 나라의 경제 규모 및 성장 속도를 보여주는 지표

GDP가 높은 국가일수록 행복할까 ?

➤ GDP가 높은 국가일수록 행복할까?



○ 프로젝트 목적

“ GDP가 높을수록 행복 지수가 높을 것이다 ”
귀무가설 설정



EDA 및 시각화



1. GDP와 행복 지수의 상관관계 확인
2. 그 외에 GDP와 상관관계가 있는 변수 확인

사용 데이터

➤ 사용 데이터

사용 데이터	데이터 설명	데이터 출처
GDP (current US\$)	단일 연도 공식 환율을 사용하여 변환된 국가별 GDP	The World Bank
GDP growth (annual %)	국가별 연간 GDP 성장률	
CO2 emissions (kt)	국가별 이산화탄소 배출량	
Life expectancy at birth, total (year)	국가별 출생시 기대 수명	
World Happiness Report	<ul style="list-style-type: none">- 세계 행복 상태에 대한 조사- ladder score, economic production, social support, freedom, absence of corruption, and generosity- 각 요인에 대해 세계 최저의 국가 평균과 동등한 가치를 가진 가상의 국가 보다 삶의 질을 높이는데 기여하는 정도를 추정하여 수치로 나타냄	World Happiness Report
How's Life Well Being (Current Well Being Average and Deprivation)	OECD 국가의 Well Being과 관련된 다양한 지표들을 나타낸 데이터	OECD.stat

데이터 전처리

전처리 및 데이터 통합

- World Happiness Data
- GDP(current US\$)
- GDP growth(Annual %)
- CO2 emissions
- Life expectancy at birth, total(years)

- 필요한 컬럼 추출
- 2018년 기준
(World Bank 최근 데이터 부재)

MERGE

Country Name	Regional indicator	Ladder Score	...	2018 gdp	2018 growth	2018 life	2018 co2
Afghanistan	South Asia	2.694	...	1.84E+10	1.189228	64.486	7440
Albania	Central and Eastern Europe	5.004	...	1.51E+10	4.071301	78.458	5560
Algeria	Middle East and North Africa	5.043	...	1.75E+11	1.2	76.693	151670
Argentina	Latin America and Caribbean	5.793	...	5.18E+11	-2.56535	76.52	177410
...

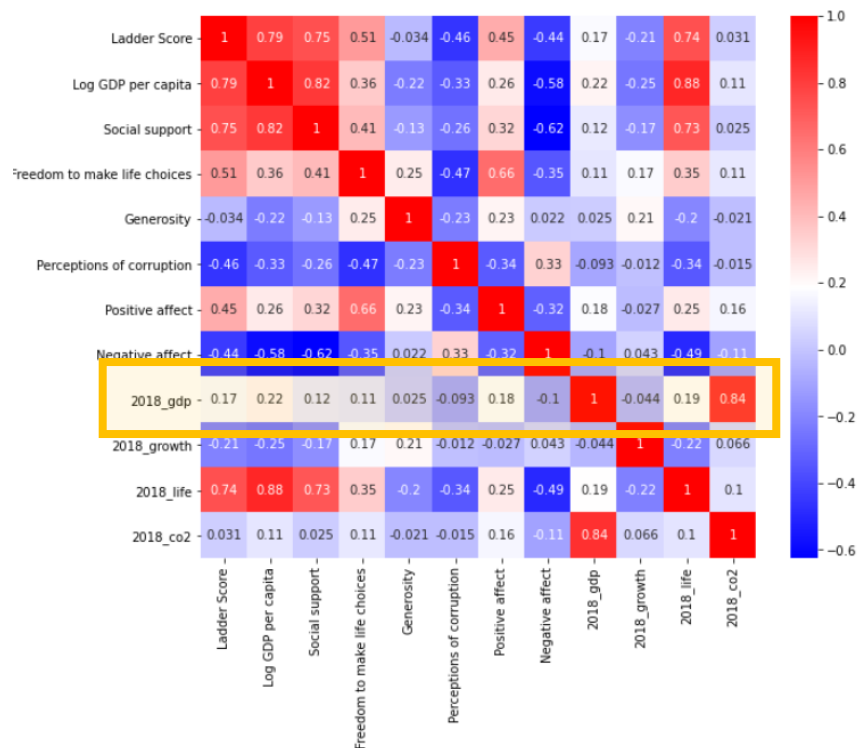
▲Total.csv

Total 이라는 이름의 데이터 프레임을 생성하여 데이터 분석 및 시각화를 용이하게 함

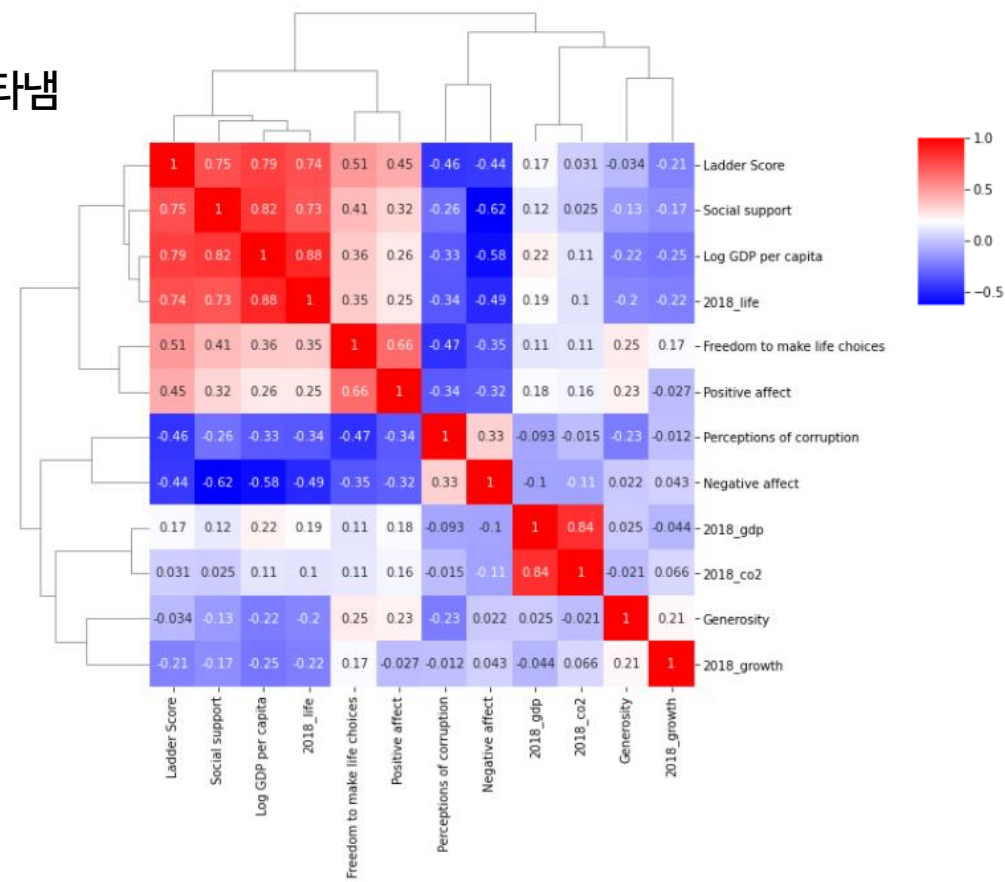
데이터 시각화

GDP는 어떤 지표와 관련이 있을까?

○ seaborn의 heatmap과 clustermap을 이용해 변수들 간의 상관관계를 나타냄



▲heatmap을 이용해 변수들 간 상관관계 나타냄



▲clustermap을 이용해 상관관계가 있는 변수끼리 묶음

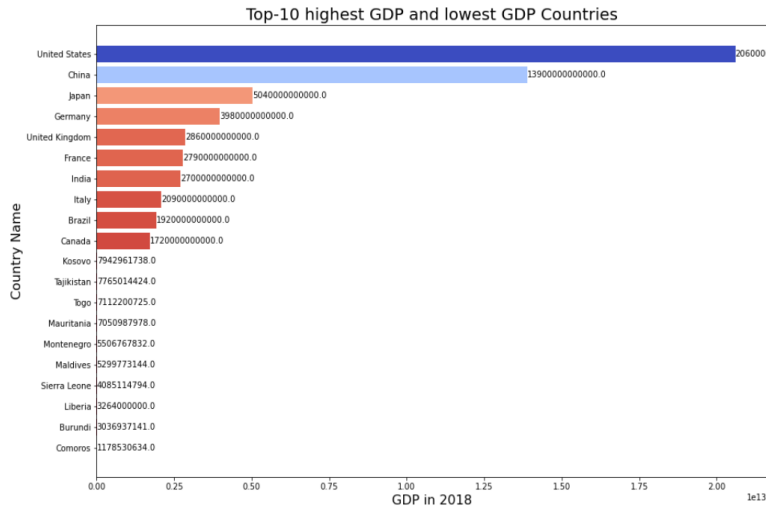
→ GDP와 CO2 emissions은 높은 양의 상관관계를 가지지만, 나머지 변수는 GDP와 유의미한 상관관계를 갖지 않음

GDP는 국가의 성장을 제대로 반영하지 못함

데이터 시각화

GDP가 국가의 행복을 반영할 수 있을까?

○ highest and lowest GDP Countries



▲ matplotlib.pyplot 이용해 시각화

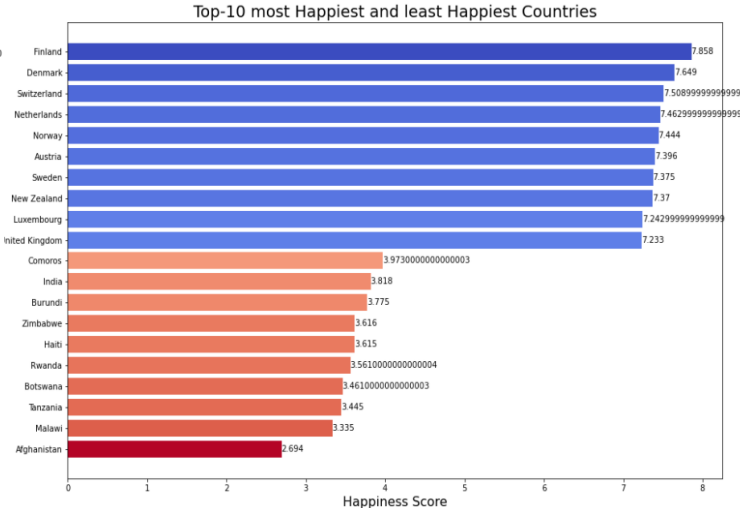
▶ 상위 Top10

: United States, China, Japan, Germany, ...

▶ 하위 Top10

: Maldives, Sierra Leone, Liberia, Burundi, ...

○ most and least Happiest Countries



▲ matplotlib.pyplot 이용해 시각화

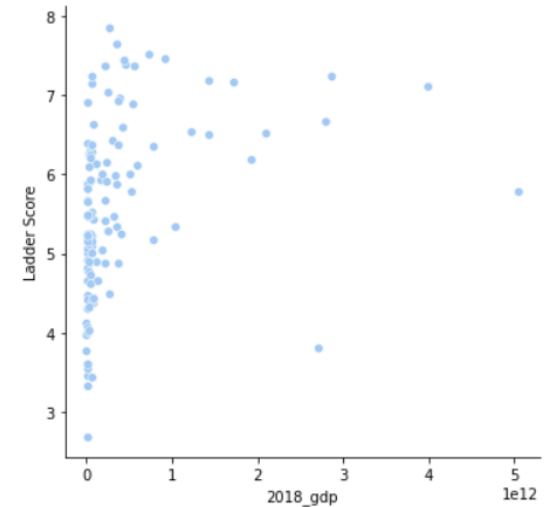
▶ 상위 Top10

: Finland, Denmark, Switzerland, Netherlands, ...

▶ 하위 Top10

: Rwanda, Botswana, Tanzania, Malawi, ...

○ GDP - 행복점수



▲ Seaborn - replot 이용해 시각화

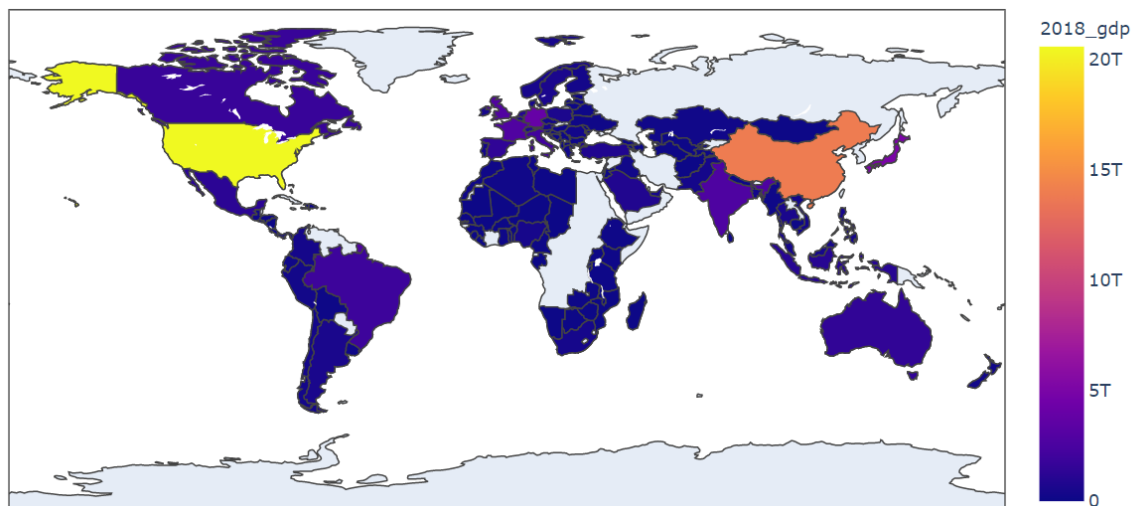
- 이상치인 미국, 중국 제외

- 규칙성을 찾을 수 없음

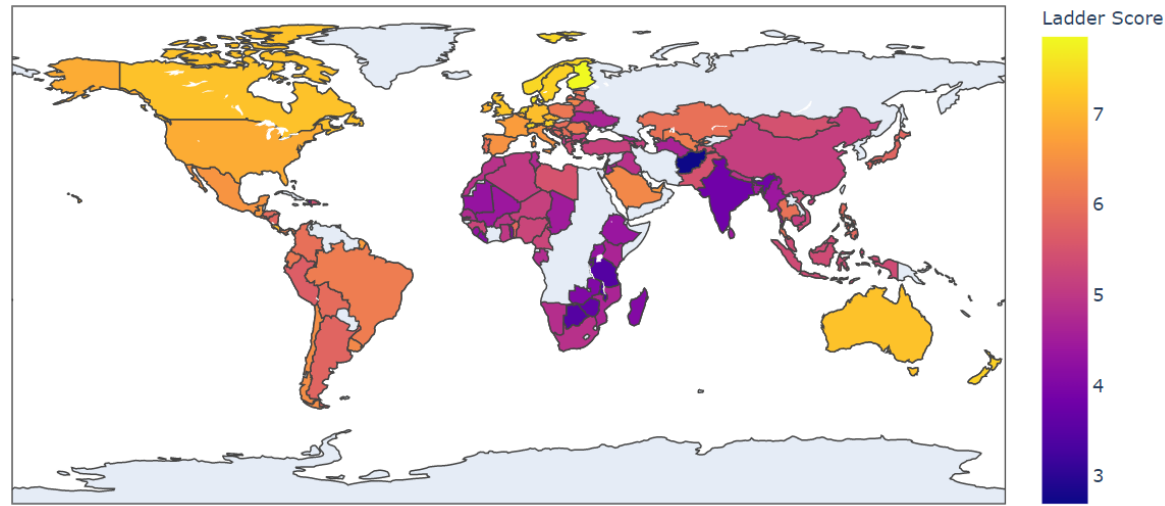
GDP와 행복점수는 상관관계가 없다 → 귀무가설을 기각한다

결과 분석

시각화 결과 분석



▲ plotly를 활용한 GDP 지도 시각화



▲ plotly를 활용한 Ladder Score(Happiness Score) 지도 시각화

- 시각화 결과에 따르면, GDP가 국민의 행복 정도에 크게 반영되지 않는 것을 볼 수 있다
- 제 6차 OECD 세계포럼에서는 경제구조의 안정성, 자산의 분배에 관한 정보를 제공하지 못하는 GDP가 경제 정책의 근간으로 사용되고 있는 것은 잘못되었다고 문제 제기함*

국가 경제가 국가의 행복지수를 나타내기 위해서는 GDP를 넘어서는 새로운 지표가 필요함

대안 제시

▶ 대안 지표 제시 : 행복지수 (Happy Planet Index)

- 영국의 신경경제재단이 2006년 7월에 도입한 지수
- 국내 총생산과 국민의 삶의 만족도, 미래에 대한 기대, 실업률, 자부심, 희망, 사랑, 지속가능성 등 무형의 기준을 합하여 인간의 행복과 삶의 질을 포괄적으로 고려하여 수치로 표현하는 지표 → 사람들이 얼마나 행복한가에 기반

행복지수 = 웰빙 x 기대수명 x 부의 평등 / 생태발자국*

*생태발자국: 인간이 자연에 살면서 남긴 영향
→ 생태 발자국이 클수록 지구에 악영향

- 직접 행복지수 지표 계산
- 행복지수 계산을 위해 사용한 데이터

행복지수 파라미터	데이터	데이터 출처
웰빙**	Current Well-being (average and deprivation)	OECD.Stat
기대수명	Life expectancy at birth, total (year)	The World Bank
부의 평등	GDP (current US\$)	
생태발자국	CO2 emissions (kt)	

**공기 오염, 약물로 인한 죽음, 수입, 고용률, 수입, 사회적 지지도 등 다양한 웰빙 수치 존재 10

대안 제시

▶ 행복 지수 계산

○ 데이터 전처리 및 통합

- ① 2010년 이후 데이터 취합
- ② 필요없는 행 삭제 및 국가와 웰빙 지표 기준 groupby
- ③ 데이터 정규화 (MinMaxScaler)
- ④ 긍정적인 지표는 더하고, 부정적인 지표는 빼서 웰빙값 생성

계산
➡

행복지수 = 웰빙 x 기대수명 x 부의 평등 / 생태발자국*

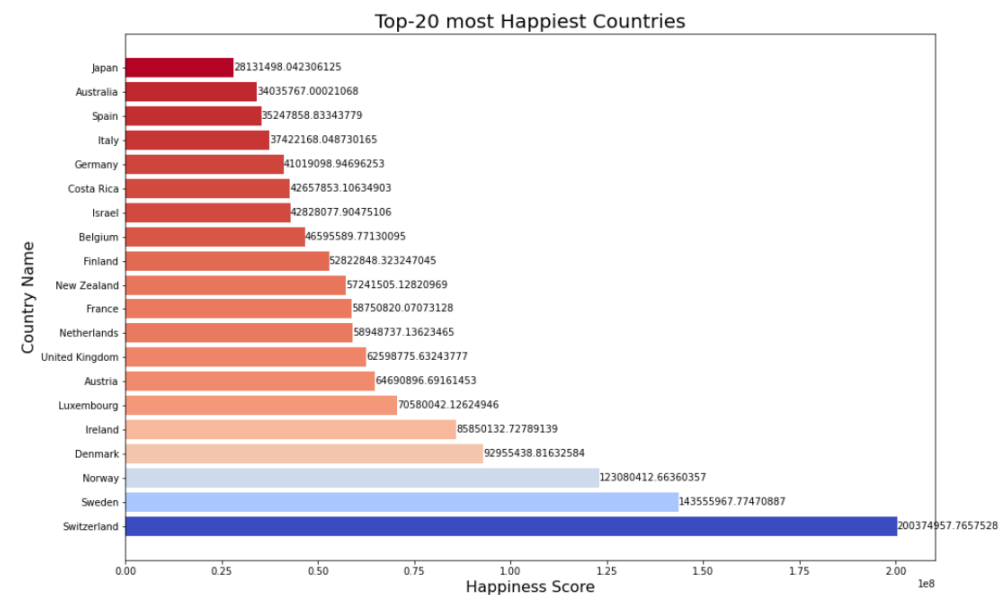
○ 행복 지수 계산 결과

Country Name	Regional indicator	Ladder Score	...	2018 gdp	2018 growth	2018 life	2018 co2	Happy Planet Index
Afghanistan	South Asia	2.694	...	1.84E+10	1.189228	64.486	7440	3.403577e+07
Albania	Central and Eastern Europe	5.004	...	1.51E+10	4.071301	78.458	5560	6.469090e+07
Algeria	Middle East and North Africa	5.043	...	1.75E+11	1.2	76.693	151670	4.659559e+07
Argentina	Latin America and Caribbean	5.793	...	5.18E+11	-2.56535	76.52	177410	2.535861e+07
...	

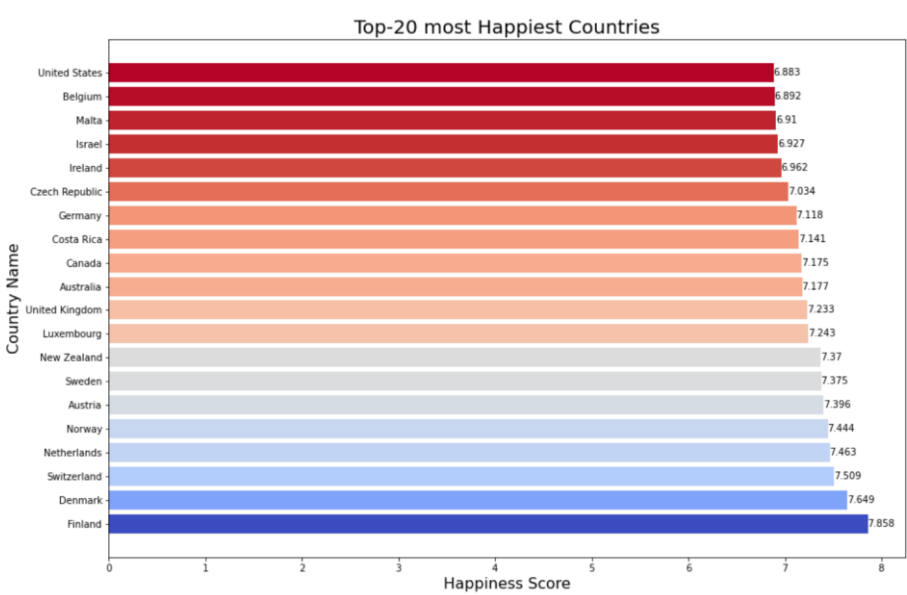
대안 제시

행복 지수 계산

○ 새로운 지표를 통해 계산한 Top-20 Most Happiest Countries



▲ matplotlib.pyplot 을 이용해 행복 지수 순위 시각화



▲ matplotlib.pyplot 을 이용해 행복 점수 순위 시각화

○ 행복 지수 TOP20와 행복 점수 TOP20의 교집합 국가

Australia, Germany, Costa Rica, Israel, Belgium, Finland, New Zealand, Netherlands, United Kingdom, Austria, Luxembourg, Ireland, Denmark, Norway, Sweden, Switzerland

→직접 구한 각 나라의 행복 지수 순위는 행복 점수 순위와 유사함 (80% 일치)

한 국가의 행복을 반영할 수 있는 지표인 행복지수를 GDP의 대안으로 제시한다

최종 결과

> 의의 및 기대효과

- GDP는 국가 경제를 나타내지만, 국가 성장 및 국민의 행복도에는 크게 반영되지 않는다는 것을 데이터 시각화를 통해 확인함
- GDP를 이용하여 국가 성장 및 행복과 유의미한 상관관계를 찾고자 데이터와 공식을 활용하여 새로운 대안을 제시하고 시각화함
- 직접 구한 행복 지수와 GDP 및 Ladder score 등 다양한 변수와 유의미한 상관관계를 찾음
- GDP로는 고려하지 못했던 국민의 삶의 질을 표현하는 새로운 대안 제시

> 한계점 및 추후 계획

- World Bank 데이터에 모든 국가가 포함되어 있지 않고 최근 년도의 데이터가 없는 경우가 있음
- 완벽한 데이터가 있다면 더 좋은 결론을 도출할 수 있을 것으로 예상
- 행복 지수 계산시 사용된 웰빙 데이터에는 OECD 국가만 포함되어있음
- 추후 OECD 외 국가 데이터를 이용해 프로젝트를 확장하고자 함
- 웰빙 지표를 설정하는 데에 주관성이 개입됨
- 더욱 면밀한 지표를 설정하고자 함

감사합니다

그 밖에 다양한 시각화는 주피터 노트북을 통해 보여드리겠습니다.