



5조 Mid Project Log

쌀세계 행복 보고서 데이터 :

한국인의 행복, 데이터로 답을 찾다.





▼ 전처리 과정 후 데이터 (유니온 용)

Updated_WHR_2018.csv <u>Updated_WHR_2019.csv</u> Updated_WHR_2020.csv ▼ DATA URL

2020~: https://worldhappiness.report/analysis/#:~:text=The World Happiness Report is, combine to affect our happiness

2015~2019: https://www.kaggle.com/datasets/sazidthe1/global-happinessscores-and-factors?select=WHR_2020.csv

▼ 원본 데이터

WHR2018.csv

WHR2019.csv

Updated_WHR_2021.csv

Updated_WHR_2022.csv

▼ 하나의 데이터셋으로 통합한 데이터

WHR_2018_2022.csv

WHR2020.csv

WHR_2021.csv

WHR_2022.csv

▼ 외부 데이터

OECD_Working_Hours_Annually(2018~2022).csv

OECD_Quality_of_Support_Network.csv



데이터

156개국의 주관적 행복 지표와 그 결정요인을 조사하여 국가별 행복 순위를 매 긴 공개 데이터셋입니다 (<u>Analysis</u> <u>|</u> The World Happiness Report). CSV 형식으로 여러 연도의 지표가 제공되며, 경제 규모, 사회적 지원, 기대수명 등의 변수도 포함되어 있습니다.

3. 세계 행복 보고서 데이터

본석 아이디어 : 나라별 행복 점수와 GDP, 사회적 지원 등 요인 간 상관 관계를 분석하여 삶의 질에 영향을 미치는 주요 요인을 탐색하거나, 지역별로 행복도가 어떤 패턴을 보이는지 시각화해 볼 수 있습니다.



변수 구분	WHR	변수 설명	⊖변수 담당
국가	country	해당 데이터가 속하는 국가의 이름을 나타냄 데이터 분석에서 국가별 행복 지수를 비교하고, 국가간 차이 를 분석할 때 사용됨	
지역	region	해당 국가가 속한 대륙 또는 지역을 나타내는 변수 지역별 정책 차이나 문화적 요인이 행복도에 미치는 영향을 비교할 때 중요한 변수로 작용	
행복 점수	happiness_score	해당 국가의 전반적인 행복 수준을 수치화한 값 점수가 높을수록 해당 국가의 전반적인 행복도가 높음을 의 미	Target 변수
경제 수준 (GDP)	gdp_per_capita	국가의 경제적 풍요 정도를 나타내는 지표로, 1인당 국내총생 산을 의미 경제적 여건이 행복에 미치는 영향을 평가할 때 사용	최혜은
사회 지원	social_support	국민이 위기 상황에서 받을 수 있는 사회적 지원의 정도를 나타냄 사회적 연결망과 신뢰 수준을 반영하는 변수	박선우
건강한 기대 수명	healthy_life_ expectancy	국민이 건강하게 기대할 수 있는 평균 수명을 나타냄 건강 상태가 행복도에 미치는 영향을 분석할 때 중요한 요소	홍다영
삶의 선택 자유	freedom_to_make_life_choices	개인이 자신의 삶에 대해 선택을 할 수 있는 자유의 정도를 평가하는 변수	
관대함	generosity	해당 사회 구성원들이 얼마나 관대하고 서로 도와주는지를 나타내는 지표	
부패 인식	perceptions_of_ corruption	정부나 공공기관에 대한 부패 인식 정도를 나타내며, 사회적 신뢰와 관련된 요인으로 작용	강선우
외부 데이터	- 교육 수준 - 전쟁의 영향 - 도시화 수준 - 기후 변화	근거를 뒷받침할만한 추가 데이터를 가져 올 수 있다.	

✓ 가설 설정

- ▼ 데이터 전반적인 가설 설정 (거시적인 가설 설정)
 - 행복 지수는 특정 소수 요인보다는 다중 요인의 영향을 받을 것이다.
 - 。 개별 변수(GDP, 건강, 사회적 지원 등)보다 복합적인 요인들이 작용할 가능성이 높음.

- 。 다중 회귀 분석을 통해 행복 지수에 대한 주요 결정 요인을 도출할 필요가 있음.
- 행복도가 높은 국가들은 공통된 정책적·사회적 패턴을 가질 것이다.
 - 。 상위 행복 국가들의 정책, 사회 구조, 경제 모델을 비교 분석하여 **행복을 높이는 공통 요소**를 찾을 수 있음.
- 행복 지수는 전 세계적으로 장기적인 상승 혹은 하락하는 패턴이 존재할 것이다.
 - \circ 예를 들어, 글로벌 경제 위기, 팬데믹, 기후변화, 전쟁 등 거시적 요인이 행복도 변화에 영향을 미칠 가능성이 있음.
 - 。 시계열 분석을 통해 장기적인 추세를 파악할 필요가 있음.
- 행복도는 문화적 요인에 의해 영향을 받을 가능성이 있다.
 - 。 종교, 가족 구조, 공동체 의식 등과 같은 **비경제적 요인**이 행복도에 영향을 미칠 가능성이 있음.
 - 。 예를 들어, **공동체 중심의 문화(아시아, 남미) vs 개인주의 중심 문화(서구권)**에 따른 행복도의 차이를 분석할 필요가 있음.
- 자연재해, 기후 변화 및 지정학적 갈등이 행복도에 부정적인 영향을 미칠 것이다.
 - 。 최근 **기후 변화 및 환경 문제**가 국가 행복도에 영향을 주는지 분석할 필요가 있음.
 - 。 예를 들어, 기후 변화의 영향을 많이 받는 국가(저지대 국가, 극단적 기후 지역)에서 행복 지수가 낮을 가능성이 있음.

▼ 변수별 가설 설정

변수 구분	WHR	가설
국가	country	- 특정 국가의 경제 및 사회적 정책이 행복도에 영향을 미칠 것이다.
지역	region	- 지역별로 행복 지수의 평균이 다를 것이다. - 선진국이 많은 지역일수록 행복 지수가 높을 것이다. - 특정 문화권(북유럽, 동아시아)이 다른 문화권보다 행복 지수가 높을(낮을)것이다.
행복 점수	happiness_score	- 행복 지수는 다양한 경제적, 사회적 요인에 의해 결정된다. - 시간에 따라 행복 지수의 평균이 변화할 것이다.
경제 수준 (GDP)	gdp_per_capita	- 1인당 GDP가 높을수록 행복 지수도 증가할 것이다. - 경제 성장률이 높은 국가일수록 행복 지수가 증가할 것이다. - 소득 불평등이 심한 국가일수록 행복 지수가 낮을 것이다.
사회 지원	social_support	- 사회적 지원이 높은 국가일수록 행복 지수가 높을 것이다. - 실업률이 낮은 국가일수록 사회적 지원이 강하고 행복도가 높을 것이다. - 사회적 연결(가족, 친구) 만족도가 높은 국가일수록 행복도가 높을 것이다.
건강한 기대 수명	healthy_life_expectancy	- 건강 기대 수명이 긴 국가일수록 행복 지수가 높을 것이다. - 공공 의료 시스템이 강한 국가일수록 건강 기대 수명이 길고 행복도가 높을 것이다. - 운동량이 많은 국가일수록 건강 기대 수명이 길고 행복 지수가 높을 것이다.
삶의 선택 자유	freedom_to_make_life_choices	- 개인의 자유도가 높을수록 행복 지수가 높을 것이다. - 여성의 권리가 보장된 국가일수록 행복 지수가 높을 것이다. - 언론 자유도가 높은 국가일수록 행복 지수가 높을 것이다.
관대함	generosity	- 기부 문화가 발달한 국가일수록 행복 지수가 높을 것이다. - 자원봉사 참여율이 높은 국가일수록 행복 지수가 높을 것이다. - 종교적인 국가일수록 행복 지수가 다르게 나타날 것이다.
부패 인식	perceptions_of_corruption	- 부정부패 인식이 낮은 국가일수록 행복 지수가 높을 것이다. - 정치적 투명성이 높은 국가일수록 국민들이 행복할 것이다. - 기업 투명성이 높은 국가일수록 국민 행복도가 높을 것이다.
외부 데이터	- 교육 수준 - 전쟁의 영향 - 도시화 수준 - 기후 변화	- 교육 수준이 높은 국가일수록 행복 지수가 높을 것이다. - 전쟁이나 정치적 불안정성이 높은 국가일수록 행복 지수가 낮을 것이다. - 도시화가 진행된 국가일수록 행복 지수가 높을 것이다. - 기후 변화 및 자연재해의 빈도가 높은 국가일수록 국민 행복 지수가 낮을 것이다.



💆 선정 이유 (문제 정의)

- WHR 데이터를 활용하여 행복 지수를 결정하는 주요 요인들을 파악하고, 국가별 행복 격차의 원인을 분석하며,
- 이를 통해 한국의 행복도를 높일 수 있는 정책 방향을 제안하는 것을 목표한다.



✔ 데이터를 통한 분석 목표(아이디어)

- 사람의 행복 지수에 가장 큰 영향을 주는 것과, 가장 적은 영향을 주는 것을 분석
- 태어난 환경(GDP 등)이 사람의 행복에 영향을 끼칠 수 있는지?

김현태

• 국가(지역)별 행복지수 비교 (왜이곳은 행복할까?)

- 나라별 행복지수 분석 및 행복지수 차이에 대한 원인 분석
- 24년도 기사에 따르면 우리나라의 행복지수 순위는 52위이다. 높은 경제적 발전에 비해 낮은 행복지수의 원인을 찾으려한다.

최혜은

- 코로나19를 기준으로 2018~2019 vs 2020~2022 데이터를 분석
- 높은 행복지수를 유지하는 나라의 주요 원인이 무엇인가를 알아내고자 한다.

• 추가적인 기후 데이터를 참고하여 기후가 행복지수에 영향을 끼치는지 알아내고자 한다.

홍다영

- 행복 지수와 건강 기대 수명의 가설 세우기
- 연도별 건강 기대 수명
- 연도별 건강 기대 수명과 행복 점수의 상관관계 확인
- 대륙별 평균 기대 수명과 평균 행복 점수 확인
- 코로나 19 확진자 수와 건강 기대 수명 연관 시켜보기



데이터 전처리

▼ 1. 데이터 병합을 위해 각 데이터에 연도를 추가

```
import pandas as pd
# CSV 파일 읽기
df = pd.read_csv('/content/WHR_해당연도.csv')
# 'Year' 컬럼 추가 및 모든 값을 해당연도으로 설정
df['year'] = 해당연도
```

▼ 2. 각 연도별 데이터에서 국가명이 다르게 기록된 경우를 확인하고, 하나의 국가명으로 일치 및 삭제

구체적인 전처리

- 1. 결측치 확인 및 처리 결정 (단 1개이므로, 추가적인 결측치 처리 💢
- 2. 1개년에만 존재하는 국가의 경우 비교가 불가능하다고 판단하여 삭제
- {"Angola", "Belize", "Sudan"} → 2018년 데이터만 존재
- {"Czechia"} → 2022년 데이터만 존재
- 3. 2020년 데이터를 기준으로 국가명을 일치시킨다.
- "Congo"의 경우 :

2018~2021년 데이터에서는 두 개의 국가(Congo (Brazzaville), Congo (Kinshasa))로 구분되어 있으나, 2022년 데이터에서는 하나로 통합되어 나타남.

따라서, 공식 사이트의 리포트를 확인한 후 Congo (Brazzaville)로 변경하는 전처리를 진행.

```
import pandas as pd
# 파일 경로
file_paths = {
  "2018": "/content/WHR_2018.csv",
  "2019": "/content/WHR_2019.csv",
 "2020": "/content/WHR_2020.csv",
 "2021": "/content/WHR_2021.csv",
 "2022": "/content/WHR_2022.csv",
# 데이터 로드
data = {year: pd.read_csv(path) for year, path in file_paths.items()}
# 국가명 컬럼 확인 (각 연도별)
columns_info = {year: df.columns for year, df in data.items()}
# 2022년 데이터의 컬럼명 변경
data["2022"].rename(columns={"Country": "country", "Region": "region"}, inplace=True)
data["2022"].columns
# 2020년 기준 국가 목록 가져오기
country_2020 = set(data["2020"]["country"])
# 각 연도의 국가명 비교
country_mismatches = {}
for year, df in data.items():
 if year == "2020":
    continue # 기준 데이터이므로 비교하지 않음
 mismatched_countries = set(df["country"]) - country_2020
 country_mismatches[year] = mismatched_countries
# 국가명이 일치하지 않는 경우 확인
country_mismatches
```

```
# 삭제할 국가 (1개년 데이터에만 존재하는 국가)
countries_to_remove = {"Angola", "Belize", "Czechia", "Sudan"}
# 변경할 국가명 매핑 (2020년 기준으로 일치)
country_mapping = {
 "Congo": "Congo (Brazzaville)",
"Eswatini": "Swaziland"
# 매핑 적용 및 삭제 처리
for year, df in data.items():
  # 국가명 변경
  df["country"] = df["country"].replace(country_mapping)
  # 삭제할 국가 제거
  df = df[~df["country"].isin(countries_to_remove)]
  # 수정된 데이터 반영
  data[year] = df
# 수정된 데이터를 다시 저장
updated_file_paths = {}
for year, df in data.items():
  save_path = f"/content/Updated_WHR_{year}.csv"
  df.to_csv(save_path, index=False)
  updated_file_paths[year] = save_path
```

▼ 3. 태블로에 맵파트 출력시 오류 해결



!!오류 해결

홍콩, 키프로스, 대만의 위치를 인식하지 못하므로 직접 드롭다운에서 선택하여 표시.



×

데이터 시각화 아이디어

▼ 그래프

- 1. 행복 지수(World Happiness Score)와 각 변수 간의 상관 관계
 - 산점도(Scatter Plot)로 시각화하여 주요 변수들의 관계 분석
- 2. 행복 지수의 연도별 변화 (2018~2022년)
 - Line Chart(선 그래프)를 이용하여 행복 지수가 증가하는 국가 vs 감소하는 국가 비교
- 3. 세계 지도(World Map) 시각화
 - 국가별 행복 지수를 색깔별로 표현하는 Choropleth Map

4. 행복 지수가 높은 국가 vs 낮은 국가 비교

- Bar Chart를 활용하여 행복 지수 Top 10, Bottom 10 국가 비교
- 5. 각 지역별 행복 지수 비교
 - Box Plot(박스 플롯) 혹은 평균 행복 지수 비교
- 6. 행복도 변화 예측 (Forecasting)
 - 트렌드 라인을 활용하여 향후 행복 지수 변화 예측



태블로, 코랩 수행 작업

강선우

▼ 3/15~16

<u>행복점수와부정인식지수간의관계.twb</u>

- Perceptions_Of_Corruption이 Happiness_Score에 주는 영향을 분석해봤음
- 계산필드로 상관계수를 구함
- : Corr([Perceptions Of Corruption], [Happiness Score])
- 상관계수는 0.4121 정도로 미미한 영향을 끼침
- 2018년과 2022년 간의 Happiness_Score, Perceptions_Of_Corruption의 차이를 구해보고, 절대적인 차이가 많이 나는 국가, 점수 혹은 인지도가 발전한 국가, 하락한 국가를 각각 Top20씩 확인해보았음
- 계산필드로 2018년과 2022년의 각 카테고리별 차이를 구함
 : SUM(IF YEAR([Year]) = 2022 THEN [카테고리] END) SUM(IF YEAR([Year]) = 2018 THEN [카테고리] END)
- 계산필드로 해당 차이의 절대값을 구함
 : ABS(SUM(IF YEAR([Year]) = 2022 THEN [카테고리] END) SUM(IF YEAR([Year]) = 2018 THEN [카테고리] END))
- 나눈 이후 부정인식과 행복점수 변화의 비례, 혹은 반비례 여부를 해당하는 국가들을 통해 확인해봄
- ▼ 3/17

<u>하위20개국가.twb</u>

- 행복지수 하위 20개 국가의 부패인식/관대함/삶의선택자유 3가지를 확인
- 사회부패도&자유도가 상위권국가에 비해 크게 영향을 끼치지 않음
- 이상치 : 아이티 관대함 / 르완다 부패인식
- 급격히 변화하는 구간 및 눈여겨볼 구간: 18~19년 브룬디 / 전체적으로 아프가니스탄&짐바브웨&시에라리온 하락, 코모로스는 2019년 이후 꾸준히 상승

▼ 3/18

• 하위20개국 필터링

<u>하위20개국필터.ipynb</u>

bottom_20_WHR_2018_2022.csv

• 한국만 필터

SouthKorea필터.ipynb

KOR_WHR_2018_2022.csv

▼ 3/19

• 기존 언론신뢰도 데이터셋에, 국가column이 지정되어있지 않은 관계로, 국가column을 넣어주고, 이후에 통합하는 코드

<u>언론신뢰도데이터셋.ipynb</u>

• 정부신뢰도 데이터 전처리

<u>정부신뢰도전처리 (1).ipynb</u>

김현태

- ▼ 3/15~16
- ▼ 3/17
- ▼ 3/18

▼ 3/19

태블로 시각화

코랩 행복지수 대한민국 과 상위20개국 비교

박선우

▼ 3/15~16

midproject(socialsupport분석).twbx

- social_support와 행복지수의 상관관계를 분석하고 social_support top10중 낮은 행복지수 나라 분석해보기로 했음.
- 2021 년에 social_support 큰 폭으로 낮아짐 높은 상관관계가 존재하는데 행복지수는 전반적으로 유지되는데에 비해 사회지원은 왜 낮아지고 있는가? 전체적으로 낮아지고 있는지/ 몇몇 국가에서 큰 폭으로 떨어졌는지 확인
- 2020_2021happinesschange / 2020_2021socialsupportchange 계산필드 추가해 증가와 감소 색으로 시각화
- 가설 설정 : happiness score이 낮은 곳에서 social support의 하락률이 높을 것이다. (O)

▼ 3/17

(mid_3_17)Bottom20변수분석(GDP, 사회지원, 기대수명).ipynb

- Bottom20의 GDP, 사회지원, 기대수명 변수 분석
- 상관관계와 변수별 시계열 분석 boxplot 시각화
- 모든 변수에서 bottom20이 더 낮은 값을 보임. 하지만 gdp에 이상치 보임
- gdp가 높지만 행복지수는 낮은 국가는?
- 중동 및 북아프리카 국가들 → 경제적으로 부유하지만 정치적 자유와 사회적 자유가 제한적인 경우가 많고, 부유한 계층과 빈곤한 계층 간의 불평등이 심각할 수 있습니다. 이로 인해 행복 점수가 상대적으로 낮을 수 있습니다.
- 한국과 bottom20 비교 시각화

▼ 3/18

(mid_3_18)한국과비교.ipynb

- 한국의 전체 순위 분석해 전체강점/ 약점 나눔
- 한국과 GDP 비슷한 나라들(10% 이내) 비교해 한국이 높은 GDP에 비해 왜 행복지수가 낮은지 분석 및 시각화
- 순위가 낮은 컬럼(사회지원, 자유도) 분석 위해OECD Better Life Index에서 한국의 자유도와 사회지원에 관련해 참고

▼ 3/19

- OECD Better Life Index에서 한국의 자유도에 관련한 지표 분석
- 한국의 행복도 증가 정책 조사

최혜은

▼ 3/15~16

- ▼ 전체 데이터 탐색 (EDA)
 - 연도별 행복 지수 변화 (라인 차트)
 - 국가별 행복 지수 분포 (히스토그램)
 - 행복 지수 요인 간 상관관계 분석 (히트맵)
- ▼ 행복 지수 ~ GDP 파악
 - gdp를 분위별로 파악하기 위해 총 3개의 그룹을 나누어 확인
 - 저소득(하위 33%), 중위권(중위 33%), 고소득(상위 33%)로 설정
 - 계산필드를 이용해 WINDOW_PERCENTILE(AVG([gdp_per_capita]), 0.33) = 0.7258
 - WINDOW_PERCENTILE(AVG([gdp_per_capita]), 0.66) = 1.113
 - gdp_group:

IF [Gdp Per Capita] < 0.7258 THEN "Low gdp (bottom 33%)"

ELSEIF [Gdp Per Capita] <1.113 THEN "Mid gdp (middle 33%)" ELSE "High gdp (top 33%)" END

250316 최혜은.twbx

▼ 3/17

<u>250317 상관행렬, 한국과의비교</u>

- 모든 변수, 행복 지수 상관관계 분석
- 상위 20개 vs 한국 비교
- 외부데이터 근무시간 데이터를 가져와 비교

▼ 3/18

250317~18 한국비교 + 근무시간

- 한국과 유사한 국가를 분석 (행복지수를 기준으로 위 아래 5개씩 추출)
- GDP, 행복지수를 기준으로 군집분석 한국은 중위소득, 행복순위 중위권에 해당
- 근무시간 데이터 → 이스라엘&코스타리카 vs 한국: 근무시간은 비슷하지만 두 나라 모두 한국보다 행복 지수가 높음

▼ 3/19

- 지원 네트워크의 질 데이터 → 이스라엘&코스타리카 vs 한국 : 한국이 월등히 낮았음을 알 수 있었다.
- 지원 네트워크의 질(Quality of Support Network)란? : 개인이 어려운 상황에 처했을 때 도움을 받을 수 있는 사회적 관계망의 질적 수준을 평가하는 지표
- 도출한 그래프를 리스트화
- 발표 자료 빌드업!

▼ 3/20

- 본격적으로 데이터를 취합하고 도출한 결과물을 발표자료에 적용!
- 전체적인 발표 흐름 및 디자인 구성

홍다영

▼ 3/15~16

- 행복지수와 건강 기대 수명 상관관계 분석해보기
- 대륙별 평균 기대 수명 상의 3개와 대룩별 평균 행복 지수 상위 3개 확인해봄.
- 연도별 평균 건강 기대 수명
- 연도별 평균 행복 점수
- 국가별 평균 기대 수명 (상위 10)

<u>중간.twbx</u>

▼ 3/17

- 평균 행복 지수 상위 20개 국가 뽑기
- 상위 20개 국가와 건강 기대 수명, 삶의 선택 자유도 상관 관계 파악
- 건강 기대 수명에서 미국과 체코 이상치 발견
- 삶의 선택 자유도는 이상치 없음
- 미국과 체코와 왜 이상치 값을 갖는가?

▼ 3/18

- health expenditure and finaning (OECD) 데이터 가져오기
- 연도별 상위 20개 국가 의료비 평균 확인
- 미국, 뉴질랜드 의료비 지출이 큰 것을 확인 → 미국이 건강 기대 수명 이상치 갖는 이유 확인
- Perceived health status by socio-economic status (OECD) 데이터 가져오기
- 체코는 의료불평등이 심각 → 체코가 이상치 갖는 이유 확인
- 뉴질랜드는 건강 상태 인식이 높아 의료비 지출이 컸던 것을 확인
- https://colab.research.google.com/drive/15aUuOWhLn4tJy0UZ-gy2ldoJzA2t8-EJ?usp=sharing

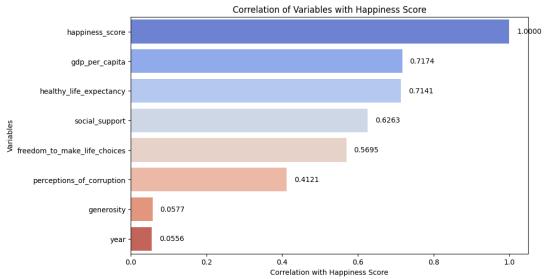
▼ 3/19

- Alcohol consumption (OECD) 데이터 가져오기
- 체코 알코올 소비량 높은 거 확인
- Body Weight (OECD) 데이터 가져오기
- 미국, 뉴질랜드 비만율 높음 → 뉴질랜드 의료비 지출이 컸던 이유 확인
- 미국 : 비만율이 높고 의료비가 비싸 건강기대수명 낮음
- 체코: 알코올 소비량이 많고, 의료불평등이 커서 건강기대수명 낮음
- 뉴질랜드 : 비만율이 높아 의료비 지출이 크지만, 의료불평등이 낮아서 건강기대수명이 높음
- https://colab.research.google.com/drive/15aUuOWhLn4tJy0UZ-gy2ldoJzA2t8-EJ?usp=sharing

▼ 상관행렬 분석 및 시각화

<u>[5조 상관관계 분석</u>





❤ 발표 자료

- https://www.canva.com/design/DAGiJbCiLDM/Fg_baWgx4J5eJ0cJVIMCuQ/edit
- 제목 설정 : 한국의 행복지수 상승 전략: 세계 행복 보고서를 통한 분석
- 목표 설정
 - 。 상위 20
 - 。 하위 20
 - 。 한국의 정책 수립
- 각각의 결과물 → 한국에 적용

💅 도한 시각화 리스트

▼ 강선우

• 딱히 () 표기 없으면 막대그래프입니다

- 5개년 평균 행복지수(맵)
- 부패인식지수(맵)
- 부패인식지수와 행복지수 간의 상관관계(산점도)
- 국가 별 평균 행복지수와 부패인식지수, 그리고 상관관계(표)
- 2018년과 2022년 간의 행복지수차(막대그래프)
- 2018년과 2022년 간의 행복지수가 가장 많이 차이나는 국가 top20
- 가장 불행해진/행복해진 국가 각각 TOP 20
- 2018년과 2022년 간의 부패인식지수가 가장 많이 차이나는 국가 20
- 부패인식지수가 가장 많이 하락/상승한 국가 20
- 부패인식지수 변화가 가장 크고, 행복지수 변화가 가장 큰 국가들
- 가장 행복해졌으며 부패인식이 높아진 국가 / 가장불행해졌으며 부패인식이 낮아진 국 가

[하위20개국+한국]

- 행복지수 하위 20개국
- 행복지수와 부패인식지수/관대함/자유도 간의 상관계수(표)
 - 。 하위권 20국가만 따로 계산
 - 。 하위권 20국가 각각 관계지수
- 하위권 20국가의 5년간 행복지수 변화(선그래프)
 - 。 부패인식지수/관대함/자유도 선그래프
 - 。 박스플롯을 통한 이상치 확인
- 하위20개 국가 관대함/자유도/부패인식지수와 행복도 간의 관계 (산점도)
- 하위20개 국가 평균 관대함/자유도/부패인식지수(표)
- 하위권 국가와 한국 행복지수 비교 및 하위권국가 평균과 비교
 - 부패인식은 하위권 국가에 비해 높고, 관대함은 확연할 정도로 낮으며, 자유도는 생 각한 것보다 낮은 편에 속했음
- > 부패인식이 강한 영향을 끼친다는 생각을 해서, 언론신뢰도/정부신뢰도를 확인해야 할 것 같다고 생각함

[Top20이면서OECD인 국가와 한국 비교(외부자료)]

- 행복지수 top20국가 중 OECD에 해당하는 국가들과 언론신뢰도/정부신뢰도 비교(선)
- 정부신뢰도/언론신뢰도와 행복지수 관계(산점도)
- 언론신뢰도와 정부신뢰도 관계(산점도)
- 부패인식과 정부/언론 신뢰도 관계(산점도)
- 행복지수 top20국가 중 OECD 해당 국가들과 평균 언론신뢰도/정부신뢰도/행복지수 비교 그래프 - 한국이 확연하게 셋 다 최하에 위치해있었음(표)
- 정부신뢰도, 언론신뢰도, 행복지수 각각 상관계수(표) 정부신뢰도-행복지수가 상관관 계 가장 높았음

정부신뢰도를 높이면 한국의 행복지수를 높일 수 있을 것

(시간이 된다면) 혜은님이 주신 자료로 정부신뢰도가 높은 국가와 한국을 비교해 개선방안 제시할것?

▼ 김현태

태블로

- 상위20개국: 행복지수맵,테이블순위,막대그래프
- 행복지수 상위20개국과 관대한(평균) 비교 다중그래프
- 행복지수 상위 20개국과 부패인식(평균) 비교 다중그래프
- 행복지수(평균)와 관대함(평균) 상관관계 그래프
- 행복지수(평균)와 부패인식(평균) 상관관계 그래프
- 년도별 대한민국과 행복지수(평균) 관대함(평균)과 행복지수 다중그래프
- 년토별 대한민국과 부패인식(평균)과 행복지수(평균) 다중그래프

코랩

- 행복지수 상위20개국과 대한민국 행복지수 비교 막대그래프
- 행복지수 상위 20개국과 대한민국 관대함 비교 막대 그래프
- 행복지수 상위 20개국과 대한민국 부패인식 비교 산점도
- 대한민국 행복지수 위아래 5개국 관대학과 부패인식비교

▼ 박선우

<전체>

- 연도별 국가의 행복지수(맵)
- 연도별 변수 평균 전체비교(bar)
- 연도별 행복지수 평균 전체비교(bar)
- 행복도 상위 20개국
- 행복도 하위 20개국

<Social Support>

- Social Support 상위 20개국
- Social Support 하위 20개국
- 사회지원과 행복지수간의 상관관계

(p-값이 모두 < 0.0001로 통계적으로 매우 유의미)

- Social Support 상위 10개국 그외 지표 비교
- 2020년과 2021년 social_support 비교
- → 모든 나라가 감소
- 2020년과 2021년 happiness_score 비교
- Happiness Score **하위** 열개 국가의 Social Support 하락률 (표)
- Happiness Score 상위 열개 국가의 Social Support 하락률 (표)
- → 하위 열개국은 평균에 못미치고 상위 열개국은 평균.

보다 하락폭이 작은 것을 볼 수 있었다.

- 사회지원, GDP ,기대수명 상관관계 (히트맵)
- → 하위 열개국은 평균에 못미치고 상위 열개국은 평균.

보다 하락폭이 작은 것을 볼 수 있었다.

<BOTTOM 20>

- GDP 시계열 그래프
- 사회지원 시계열 그래프
- → Social Support"2020-2021년" 급격히 감소
- : tableau 에서 모든 나라가 감소한 것 볼 수 있음
- 기대수명 시계열 그래프 "2020-2021년" 급격히 감소
- 사회지원, GDP ,기대수명 하위권 / 상위권 Boxplot

(모든 변수에서 bottom20이 더 낮은 값을 보임.

하지만 gdp에 이상치 보임

- → gdp가 높지만 행복지수는 낮은 국가는?)
- Bottom 20 중 GDP 와 행복지수 관계 산점도
- → 중동 및 북아프리카 국가들 :

경제적으로 부유하지만 정치적 자유와 사회적 자유가 제한적인 경우가 많고,

부유한 계층과 빈곤한 계층 간의 불평등이 심각

<순위 데이터>

사회지원 기준

- social_support 기준 상위 10개 국의 happiness_score 순위
- social_support 기준 하위 10개 국 happiness_score 순위

<한국>

• 한국과 Bottom20 변수별 평균 비교

한국은 사회 지원, 경제 수준 (GDP), 건강한 기대 수명에서 훨씬 높은 수준 -> 전반적인 생활 수준이 높다 반면, 삶의 선택 자유, 관대함, 부패 인식에서는 Bottom 20 국가들과 비교할 때 상대적으로 낮은 점수를 기록

• 한국과 GDP비슷한나라들(10%이내) 한국과 변수 차이 시각화(bar)

41개국 중 한국의 사회 지원 (40위), 한국의 삶의 선택 자유(39위)로 두가지 지표가 낮은 순위

한국의 사회 지원 (40위), 한국의 삶의 선택 자유(39위)

보기 위해 외부 데이터 (OECD Better Life Index)

<한국의 순위 데이터>

- 총 데이터에서 한국의 변수별 순위(강점3개 약점 3개 분석)
- 한국과 GDP비슷한나라들(10%이내) 중 한국의 변수별 순위

<외부데이터 분석>

- 사회 지원 part
- 1. 노동시장의 불안정성 ⇒ 나라 수: 35 / 한국의 노동 시장 불안정성 순위: 24 (안정적)

상위권과 비교했을 때 개선의 여지

2. (시각화는 아님) 노동시장 불안정성 상위권과 비교/ 안정성 top10의 요인 분석

한국이 강화하면 좋은 것 같은 점

- 1) 실업보험 & 사회보장 강화 (독일, 룩셈부르크, 벨기에)
- 2) 탄력적 근무제 도입 확대 (네덜란드, 핀란드)

- 3) 직업 훈련 & 재교육 제도 강화 (독일, 일본)
- 4) 노사 협력 문화 정착 (핀란드, 독일)
- -> 4개의 시사점 정리
- 3. 사회적 네트워크 Total 비교 (bar) 42개국 중 40번째
- 4. (순위 데이터) 사회적 네트워크 소득 수준과 / 여성 에 따라 비교

가석 석정

- 고소득과 저소득층간에 obs_value간 격차가 클것이다. 즉, 소득이 많을 수록 사회적 지원이 더 활발하게 이루어질 것이다.
- 한국은 전체적으로 사회적 네트워크 시스템이 좋지 않다.
- 5. socialnetwork 한국저소득층, 한국고소득층, OECD 평균 , 한국total 비교(bar)
- 6. socialnetwork 남녀 비교 (bar)
- 7. socialnetwork Top10 과 한국 비교
- 삶의 자유도 파트
- 1. 여가 및 개인시간/ 투표율 모두 양호한 순위

<2024년>

2023년에는 데이터를 분석하는 방식이 달라 합치지 못했다.

2024년거만 보고 비교할 수 있어 2024년에 우리나라의 행복지수가 증가한것을 보아

2024년을 기준으로 어떤게 발전하면 좋은지 보았다

5개년데이터 -> 2024년의 결과와 비교 -> 우리의 결과

- happiness_score 2018년부터 2024년의 시계열 데이터 (2023제외)
- 한국의 2018년부터 2024년의 시계열 데이터 (2023포함)
- (18-24) 모든 나라 평균 연도별 상관분석
- : social_support가 상관관계가 제일 높음에도 떨어졌음 / 사회적 지원을 조금 더 높여서 happiness_score를 높이는 방안
- 변수별 2018년부터 2024년의 시계열

강점

- 1. 건강한 기대수명 (7위)
- 한국은 의료 서비스의 발달로 기대수명이 긴나라
- 2. GDP 순위 (29위 경제력은 높은 편
- 3. 행복점수 (59위)
- 중위권 (경제력과 비교했을 때 상대적으로 낮음)

약점

- 1. 삶의 선택 자유 (140위)
 - 사회적 압박이나 경쟁문화가 원인
- 2. 사회지원 (95위)
- 복지 살펴보기 3. 관대함 (87위)
- 4. 부패인식(79위)
 - 정부나 기업의 부패에 대한 부정적인 인식

▼ 최혜은

[전체]

- 연도별 행복 지수 변화 (라인)
- 5개년 평균 행복지수 상위(하위) 20개 나라 (막대)
- 전체 변수 / 행복 지수 상관행렬
- 연도별 행복지수 ~ GDP 상관관계
- 소득 분위별 행복지수 (boxplot)
- 지역별 행복지수와 GDP의 상관관계 (boxplot)
- 18~19 / 20~22년 행복지수 상위(하위) 20개 나라 및 GDP 수준 (막대)
- Top 20 시계열 그래프

[한국]

- 한국 연도별 변화 (라인)
- 21~22 한국의 변수 변화 (막대)
- 한국 삶의 자유도 + 행복지수 변화 (라인)
- 한국과 행복지수가 유사한 국가 10개 산출 후 산점도
- 근접국가 vs 한국 변수 비교 (막대)
- 전체국가 군집분석 (0,1,2)

[근무시간]

- 근무시간을 포함한 전체 변수 상관행렬
- 근무시간 ~ 행복지수 산점도 및 상관계수
- 연도별 근무시간 행복지수 비교 (라인)
- Top20+한국 근무시간 비교 (막대)
- 근무시간 ~ 행복지수 Top20+한국 산점도

[지원 네트워크의 질]

- 네트워크 ~ 행복지수 산점도 및 상관계수
- 네트워크 ~ 행복지수 Top20+한국 산점도

▼ 홍다영

<전체 국가>

- 연도별 건강기대수명과 행복지수의 상관관계(산점도)
- 연도별 건강기대수명 / 연도별 행복지수 (꺾은선그래프)
- 국가별 평균 기대수명 상위10 (막대그래프)
- 대륙별 평균 기대수명 상위3 (막대그래프)
- 대륙별 평균 행복지수 상위3 (막대그래프)
- 평균 기대수명 / 평균 행복지수 (맵)

<상위 20-행복지수>

- 행복지수 평균 상위 20개 국가 (막대그래프)
- 상위 20, 하위 20 국가 건강기대수명과 삶의 선택 자유도 평균 (막대그래프)
- -> 큰 차이를 보임. 즉, 중요한 요소

<건강기대수명>

- 건강기대수명과 행복지수 상관관계 (산점도)
- -> 건강기대수명 👔 행복지수 👔
- 건강기대수명 이상치 (산점도)
- -> 미국과 체코 이상치

가설 : 미국과 체코는 의료 시스템 불평등이 있을 것이다.

미국 - 민간 의료 중심으로 의료비 🚹+ <GDP 기준 국가별 의료비 지출>

- 비만율 🚹+ <국가별 비만율>

체코 - 세계에서 맥주 소비량 🚹 + <알코올 소비량>

- 의료 불평등 👔 + <사회경제적 지위에 따른 건강 상태 인식-소득>

<삶의 선택 자유도>

- 삶의선택자유도와 행복지수 상관관계 (산점도)
- -> 삶의선택자유도 🚹 행복지수 🚹
- 삶의 선택 자유도 이상치 (산점도)
- -> 이상치 없음

가설 : 삶의 선택 자유는 행복지수에 큰 영향을 줄 것이다.

- 삶의 선택 자유도 이상치 하위 20 국가 (박스플롯)
- -> 이상치 없음 (가설 증명)

<외부데이터>

:2018-2022 기준으로 수집

<GDP 기준 국가별 의료비 지출>

- 연도별 상위 20 국가 의료비 평균 (꺾은선그래프)
- -> 코로나 19로 2020-2021 급증
- 국가별 의료비 지출 평균 그래프 (꺾은선그래프)
- -> 미국과 뉴질랜드 의료비 지출 👔 (미국 의료비 비쌈 가설 증명)

가설 : 뉴질랜드 의료비 지출이 큰 이유가 있을 것이다.

<국가별 비만율>

- 국가별 비만율 평균 그래프 (막대그래프)
- -> 미국과 뉴질랜드 비만율 🚹 (미국, 뉴질랜드 의료비 지출 큰 가설 증명)

<사회경제적 지위에 따른 건강 상태 인식-소득 하위 20%그룹>

- 국가별 사회경제적 지위에 따른 건강 상태 인식 (막대그래프)
- -> 체코는 사회경제적 지위가 작을수록 건강상태인식 낮음 (체코 의료불평등 심화 가설 증명)
- -> 뉴질랜드는 사회경제적 지위가 높을수록 건강상태인식 높음 (뉴질랜드 의료불평등 낮음 가설 증명)

<알코올 소비량>

- 국가별 알코올 소비량 (막대그래프)
- -> 체코가 알코올 소비량이 높음 (가설 증명)

<결론:

미국 : 비만율 🚹 의료비 🚹 -> 건강기대수명 ህ

체코 : 알코올 소비량 👔 의료불평등 👔 -> 건강기대수명 🚺

뉴질랜드 : 비만율 🚹 의료불평등 🚺 -> 건강기대수명 👔

<한국>

의료시스템 불평등을 🕕 정책 필요

★ 참고 자료

1. OECD better life index : OECD에서 측정한 삶의 질 데이터

https://www.oecdbetterlifeindex.org/countries/korea/

- 나라별 일자리, 교육, 환경 등 전반적인 배경을 알 수 있음
- → 이를 이용해, 이상치가 뚜렷한 나라의 원인을 분석할 수 있음.

2. Our World in Data - Happiness & Life Satisfaction : 행복과 삶의 만족도에 대한 다양한 시각적 데이터 분석

https://ourworldindata.org/happiness-and-life-satisfaction

• 전세계적인 행복과 삶의 만족도에 대한 다양한 시각화 데이터

3. OECD Data Expolorer : OECD에서 제공하는 데이터 셋

https://data-explorer.oecd.org/vis?

 $\label{localized-localiz$

4. Reuters Institute Digital News Report : 로이터에서 제공하는 뉴스에 관련된 데이터셋

- 뉴스 신뢰도 등과 관련된 데이터
- 데이터셋을 국가마다 다운받아야 한다는 단점이 있음

 $\underline{\text{https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/country-and-market-data}\\$

5. OECD Trust in Government Data : OECD에서 제공하는 정부 신뢰도 데이터셋

 $\underline{\text{https://www.oecd.org/en/data/indicators/trust-in-government.html?oecdcontrol-3122613a85-var3=2022}$