

세계 행복 지수의 주요 영향요인 탐색과 국가 유형화

-프로그램 기획서-

작성일	2025.12.06	작성자	문채은, 정연수, 이윤서
프로젝트명	세계 행복 지수 데이터 시각화		
팀원	윤하은(조장), 정연수, 이윤서, 문채은, 김연주		
	역할 분담		
윤하은(조장)	발표 및 ppt 제작		
정연수	코랩 기반 전체적인 코드 작성, 그래프 해석 도출 도움, 기획서 수정		
이윤서	주피터 환경 최종 코드 정리, 그래프 결과 해석 및 결론 도출, 기획서 수정		
문채은	전반적인 기획서 작성		
김연주	코드 해석 및 주요 개념 정리		

2. 주제 선정 및 그 이유

주제: 세계 행복 지수(*World Happiness Index*) 데이터 시각화

선정 이유: 각 국가별 세계 행복 지수 비교를 통한 객관적 삶의 질 지표에 대한 이해 필요성

세계 행복 지수(*World Happiness Index*)는 소득 수준, 건강 기대수명, 사회적 지원, 개인의 자유, 부정부패 인식 등 다양한 사회·경제적 요인을 종합하여 국가별 삶의 질을 평가하는 대표적인 국제 지표이다. 여러 요인을 복합적으로 반영한 지표라는 점에서, 이 지수는 단순한 GDP 경제력만으로는 측정하기 어려운 “삶의 만족”, “사회적 안정감”, “공동체 의식” 등 질적 요소를 평가하는 유용한 도구라고 할 수 있다.

이를 데이터로 시각화하고 분석하는 작업은, 삶의 목표로서 행복을 추구하는 현대인들에 게 ‘어떤 삶이 행복한 것인지’를 다시 생각하게 만들 수 있는 의미있는 분석이다. 이러한 탐구를 국제적 차원에서 비교, 해석할 때 그 가치와 통찰은 더 깊어질 것이다. 따라서 본 프로젝트는 특정 국가만을 집중적으로 분석하기보다는, 국제 데이터를 기반으로 행복이라는 보편적인 주제를 다각도로 바라보는 것에 목적이 있다. 세계 행복 지수를 시각화 함으로써, 국가 간 행복도 차이를 구조적으로 비교할 수 있고, 행복을 구성하는 각 요소(소득·사회적 지원·건강·자유·관용 등)가 어떤 패턴을 보이는지 확인할 수 있으며, 국가별 문화·사회적 환경이 행복에 미치는 영향을 거시적 관점에서 이해할 수 있다.

이러한 시각화 결과는 단순한 순위 나열을 넘어, 인간의 삶의 질과 행복의 본질적 요인이 무엇인지 탐구하는 데 도움을 준다. 다양한 국가를 동일한 기준으로 비교하기 때문에 분석의 객관성이 높고, 시각적 자료를 통해 복잡한 데이터를 보다 쉽게 해석할 수 있다는 장점 또한 존재한다. 이처럼 세계 행복 지수를 중심으로 한 데이터 시각화는 사회적 의미와 학문적 가치 모두를 갖춘 주제로, 국제적 관점에서 “행복”이라는 개념을 이해하는 데 유의미한 접근이다.

3. 프로그램 소개

3-1. 해결하려는 문제

세계 행복 지수는 국가 단위 삶의 질을 종합적으로 평가하는 중요한 지표이지만, 다음과 같은 한계를 지닌다.

- ◆ 단순 순위나 점수만으로는 국가 간 차이의 구조적인 배경을 이해하기 어렵다.
- ◆ 행복 지수의 영향 요인들 간 상호작용이나 관계를 직관적으로 파악하기 어렵다.
- ◆ 국가들을 유형별로 비교하거나 군집화해 분석할 수 있는 체계적 도구가 부족하다.

이 프로그램은 이러한 한계를 해결하기 위해 데이터 시각화, 통계 분석, 군집 분석을 결합하여 세계 행복 지수 데이터를 구조적으로 이해하고 비교가 가능하도록 설계되었다.

3-2. 제공하는 가치

이 프로그램은 단순 시각화를 넘어 다음과 같은 분석적 가치를 제공한다.

1) 구조적 이해 제공

행복 점수를 구성하는 요소들의 상호작용을 분석하여, 각 국가들의 행복 수준의 원인을 구조적으로 파악할 수 있다.

2) 객관적 비교 가능

행복 지수와 기여 요인을 시각적으로 비교하여 단순 순위 이상의 국가 간 차이, 특성, 패턴을 파악할 수 있다.

3) 국가 유형화

Top3 변수를 기반으로 K-means 군집분석을 수행하여 국가를 유형별로 분류함으로써, 각 국가들의 정책적·사회적·경제적 특성을 이해할 수 있다.

3-3. 전체 기능 흐름

1) 데이터 로드 및 정제

엑셀 데이터를 불러와 Data Frame으로 변환한다.

컬럼명 통일, 불필요한 변수 제거, 결측치 처리 등 분석 가능한 형태로 변환한다.

2) 데이터 탐색

상위 5개 국가들의 점수와 기여도를 확인하여 주요 지표들의 의미를 해석한다.

이와 같은 북유럽 국가들이 높은 행복 점수를 보이는 이유를 분석한다.

3) 행복 요인 간 구조 분석

상관계수를 Heatmap으로 시각화하여 GDP·사회적 지지·기대수명 간 강한 양의 상관 구조가 있음을 확인한다.

4) 주요 변수 Top3의 영향도 분석

Boxplot으로 High, low 그룹 간 행복 수준을 비교한다.

산점도, 회귀선을 통해 변수별 행복 지수와의 관계를 시각적으로 확인한다.

OLS 회귀분석으로 요인의 영향력 크기와 유의성을 검증해 어떤 요인이 행복을 가장 많이 설명하는지 통계적으로 분석한다.

5) 국가 간 군집분석

세계 지도 베이스으로 국가별 군집 결과를 시각화한다.

레이아웃 차트로 군집 Cluster0~3 지표를 비교해 국가들의 유형별 특성을 도출한다.

6) PCA 기반 국가 분포도

주요 변수인 세 변수(GDP·사회적 지지·기대수명)를 2차원(PC1·PC2)으로 축소한다. 국가 간 거리와 위치 관계를 직관적으로 시각화하여 행복 기반 자원의 상대적 격차를 파악한다.

3-4. 사용자가 얻게 되는 경험

사용자는 이 프로그램을 통해 탐색 -> 이해 -> 해석 -> 통찰 도출에 이르는 전 과정을 직접 경험하게 된다.

데이터 구조를 직관적으로 파악하고, 행복을 구성하는 주요 요인의 관계를 스스로 발견하게 된다. 국가별 행복 점수의 차이를 시각적으로 비교하여 “왜 특정 국가가 행복한가?”를 논리적으로 설명할 수 있다. 또한, 군집분석과 PCA 시각화를 통해 세계 국가들이 어떤 유형으로 분류되는지 확인하고, 국가 간 상대적 위치와 행복 기반 요인의 분포를 시각적으로 이해하는 경험을 하게 된다.

3-5. 특징 요약

이 프로그램은 세계 행복 지수를 정제·분석·시각화하여 국가 간 행복 구조를 이해하도록 설계된 통합 분석 프로그램이다. 주요 행복 요인의 영향력 분석부터 국가 유형화(K-means), PCA 기반 분포 분석까지 단계적으로 수행하며, 단순 수치가 아니라 “행복을 결정하는 구조”를 깊이 있게 해석할 수 있다.

4. 주안점

1. 리스트, 튜플, 딕셔너리

1-1 리스트 (ex. 대괄호 [] 사용, 리스트 인덱싱, 슬라이싱, 리스트 메소드 등)

TOP3 변수 기반 국가 군집분석 (세계 지도 베블맵)

```
vars_top3 = ["GDP per capita", "Social support", "Life expectancy"] → 리스트 사용
```

1-2 딕셔너리 (ex. {} 사용, 딕셔너리 메소드 등)

TOP3 변수 기반 국가 군집분석 (세계 지도 베블맵)

```
hover_data={ ... } → 딕셔너리 사용
```

2. 조건문, 반복문

2-1 조건문 (ex. 기본 if 문, if-else 문, 산술연산, 논리연산 조건문, if-elif-else 문 등)

TOP3 변수의 영향력 검증 - OLS분석

```
if matched_cols: top3_vars.append(matched_cols[0]) → 기본 if 문
```

2-2 반복문 (ex. for문, where문)

TOP3 변수의 영향력 검증 - OLS분석

```
for var in original_vars: → for 반복문
```

```
df["label"] = np.where(df[country_col].isin(label_countries), df[country_col], "") → where 반복문
```

3. 사용자 정의 함수

3-1 def()

행복지수에 영향을 미치는 변수 - Boxplot, Scatter

```
def plot_top3_box_scatter(df, var, score="Ladder score", country_col="Country name"): → 사용자 정의 함수
```

4. 문자열 처리

```
# TOP3 변수 기반 국가 클러스터링 + Plotly 버블차트
country_col = "Country name"
score_col = "Ladder score" → 문자열 생성
df["label"] = np.where(df[country_col].isin(label_countries), df[country_col], "") → 문자열 비교
```

5. 외부 파일 입출력 또는 시각화

5-1 외부 파일 입출력

```
# 세계행복지수 엑셀파일 불러오기
happiness_data = pd.read_excel("/Users/iyunseo/Desktop/세계행복지수_2025.xlsx")
시각화: matplotlib, seaborn, plotly.express, Heatmap 등 다양한 시각화 코드 사용
```

5-2 시각화

```
# 행복지수에 영향을 주는 요소 - 상관계수 히트맵
import matplotlib.pyplot as plt
import seaborn as sns
sns.heatmap(corrmat, annot=True, cmap="PiYG", center=0, linewidths=0.5)
# 국가별 군집 분석 (레이더 차트 + 국가 목록)
import plotly.graph_objects as go
```

5. 주요 기능

주요 기능/분석 기능	기대 효과	활용 분야
연도별 국가 행복 점수 시각화 (Matplotlib/Seaborn)	특정 국가의 시간에 따른 행복 점수 변화를 직관적으로 파악 가능	국가별 정책 평가, 장기적 삶의 질 변화 분석
요인별 지표 비교(Scatter, Seaborn/Matplotlib)	GDP, 사회적 지원, 기대 수명 등 주요 지표 간 관계와 패턴을 분석 가능	변수별 영향 분석, 행복 구성 요인 이해
국가 간 행복도 비교 (멀티 국가 시각화, Plotly Express)	여러 국가의 행복 점수를 동시에 비교하여 상대적 위치와 차이를 분석 가능	국제 비교, 정책 벤치마킹, 글로벌 연구
상관관계 분석 (Heatmap, Seaborn/Matplotlib)	여러 지표 간 상관관계를 색상으로 표현하여 변수 간 관계와 패턴 쉽게 확인 가능	변수 선택, 데이터 탐색, 통계 분석

인터랙티브 시각화 (Plotly Express)	마우스 hover, 확대/축소 등 인터랙티브 기능으로 데이터 탐색 효율성 향상	웹 기반 대시보드, 동적 데이터 분석, 보고서
다차원 데이터 구조 분석 (PCA)	여러 지표를 주요 성분으로 축소해 국가별 패턴과 특성 파악 가능	차원 축소, 데이터 탐색, 군집화 전 시각화
변수 영향 분석 (OLS)	지표가 행복 점수에 미치는 영향을 수치적으로 추정 가능	회귀 분석, 인과관계 분석, 예측 모델 개발
국가 클러스터링 (KMeans)	TOP3 변수 기준으로 국가를 유사 그룹으로 분류하여 구조적 패턴 분석 가능	국가/지역 비교, 비지도 학습 기반 군집화, 정책 분석

5-1 주요 기능 상세 설명 (1조_세계행복지수 코드 기반으로)

```
import pandas as pd #판다스
import matplotlib.pyplot as plt #맷플롯 라이브러리 (Matplotlib)
import seaborn as sns #씨본(seaborn)
import plotly.express as px #Plotly Express
sns.heatmap(corrmat, annot=True, cmap="PiYG", center=0, linewidths=0.5) #heatmap
from sklearn.decomposition import PCA #PCA 분석
model_clean = sm.OLS(y_clean, X_clean).fit() #OLS분석
from sklearn.cluster import KMeans #클러스터링(군집화), cluster(그룹)
```

Pandas: 표 형태의 데이터를 쉽게 다루고 분석할 수 있도록 도와주는 Python 라이브러리로, 데이터 정리, 필터링, 집계, 통계 연산, 시각화 준비 등을 효율적으로 수행할 수 있다.

Matplotlib: Python에서 가장 널리 사용되는 시각화 라이브러리이며, 데이터의 구성을 살펴볼 수 있게 하는 다양한 종류의 그래프를 지원한다.

Seaborn: Matplotlib 기반으로 더 간결한 코드로 예쁜 통계 그래프 생성할 수 있다.

Plotly Express: 몇 줄의 코드로 복잡한 그래프를 쉽게 만들고, HTML 기반의 인터랙티브 기능(줌, 클릭, 호버)을 바로 사용할 수 있는 시각화 라이브러리이다

Heatmap: 두 변수의 관계를 색상 값으로 표현하는 통계 시각화 기법으로, 통계 데이터 분석에 용이하다.

PCA: 다차원 데이터를 주요 성분으로 축소하여 차원 감소와 데이터 패턴 시각화를 돋는 통계 기법이다.

OLS: 선형 회귀 분석 방법 중 하나로, 오차 제곱합을 최소화하여 변수 간 선형 관계를 추정하는 기법이다.

KMeans: 데이터를 비슷한 특성끼리 클러스터(군집)로 나누는 비지도 학습 알고리즘으로, 중심점(centroid)을 반복적으로 업데이트하며 군집화한다.

5. 결과화면

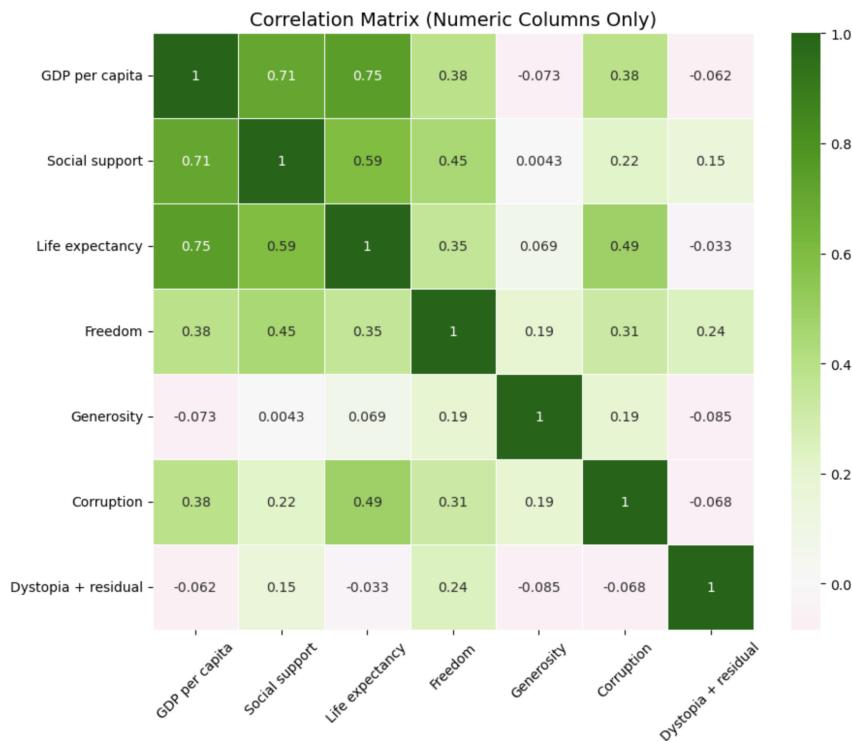
5.1 데이터 프레임의 첫 5개 행 (표 형태)

[4]:

	Country name	Ladder score	upperwhisker	lowerwhisker	Explained by: Log GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy	Explained by: Freedom to make life choices	Explained by: Generosity	Explained by: Perceptions of corruption	Dystopia + residual
0	Finland	7.736	7.810	7.662	1.749	1.783	0.824	0.986	0.110	0.502	1.782
1	Denmark	7.521	7.611	7.431	1.825	1.748	0.820	0.955	0.150	0.488	1.535
2	Iceland	7.515	7.606	7.425	1.799	1.840	0.873	0.971	0.201	0.173	1.659
3	Sweden	7.345	7.427	7.262	1.783	1.698	0.889	0.952	0.170	0.467	1.385
4	Netherlands	7.306	7.372	7.240	1.822	1.667	0.844	0.860	0.186	0.344	1.583

설명: 프로그램 실행 시 가장 먼저 세계 행복 보고서(2025)의 주요 데이터가 표 형식으로 출력된다. 이 표에는 각 국가의 행복 점수(Ladder score)와 행복 점수를 구성하는 핵심 요소 (GDP, 사회적 지지, 기대수명, 자유도, 관대함, 부패 인식 등)가 정리되어 있다. 이를 통해 전체 데이터 구조를 빠르게 이해할 수 있으며, 북유럽 국가들이 행복지수 상위권을 차지한다는 특징도 확인할 수 있다.

5.2 행복지수 기여 요소들의 상관계수 히트맵



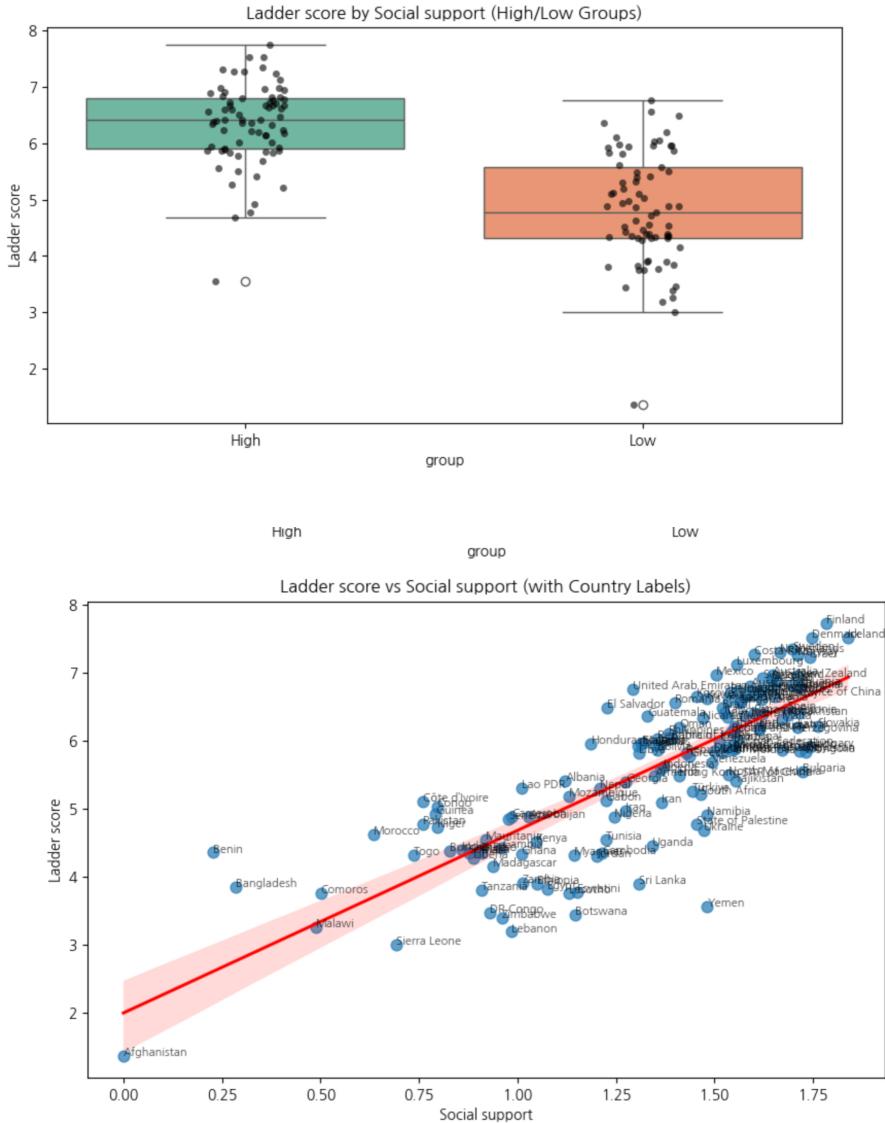
설명: 행복지수(Ladder score)와 다른 변수들의 상관관계를 히트맵으로 시각화한 화면이다.

1인당 GDP, 사회적 지지, 기대수명은 행복지수와 매우 강한 양(+)의 상관관계를 보인다.

삶의 자유(Freedom)는 핵심 요소들과 중간 정도의 양의 관계를 가지고, 관대함(Generosity)은 다른 변수들과 거의 상관관계가 없어 독립적인 문화적 요소임을 보여준다. 부패 인식 (Corruption)이 높을수록(=부패가 적을수록) 행복지수가 증가하는 경향이 있으나, 강도는 중

간 수준이다. 이를 통해 행복을 설명하는 핵심 요소를 직관적으로 확인할 수 있다.

5.3 박스플롯, 산점도, 회귀선을 활용한 Top3 변수 분석 *Top3 중 Social support만 작성



설명: 상관계수 분석을 통해 행복지수에 가장 큰 영향을 미치는 세 변수(사회적 지지, 1인당 GDP, 기대수명)를 선정하고 변수별 그래프를 출력한다.

Boxplot 결과 해석: High/Low 그룹으로 나누어 각 변수 수준이 행복지수에 어떤 차이를 만드는지 확인 가능하다.

Scatter + 회귀선 결과 해석: 점들이 회귀선 주변에 모여 있어 강한 양의 상관 관계보이며, 사회적 지지가 높아질수록 행복지수가 꾸준히 증가한다. 일부 국가가 평균 관계선에서 벗어나는 사례도 있지만 전반적으로 패턴이 명확함을 알 수 있다.

특히 세 변수 모두 행복지수와 뚜렷한 양의 상관 패턴을 보이며, 사회적 지지(Social support)가 가장 강한 영향 요인임을 알 수 있다.

5.4 회귀 분석 결과표(결측치 제거 후 / 결측치 평균 대체 후)

== 결측치 제거 후 회귀 결과 ==

OLS Regression Results

Dep. Variable:	Ladder score	R-squared:	0.759
Model:	OLS	Adj. R-squared:	0.753
Method:	Least Squares	F-statistic:	148.7
Date:	Wed, 26 Nov 2025	Prob (F-statistic):	1.28e-43
Time:	23:32:20	Log-Likelihood:	-124.65
No. Observations:	146	AIC:	257.3
Df Residuals:	142	BIC:	269.2
Df Model:	3		
Covariance Type:	nonrobust		
coef	std err	t	P> t
const	1.7632	0.193	9.128
Social support	1.7240	0.196	8.789
GDP per capita	0.6029	0.211	2.857
Life expectancy	1.3181	0.332	3.970
Omnibus:	4.588	Durbin-Watson:	1.506
Prob(Omnibus):	0.101	Jarque-Bera (JB):	4.102
Skew:	-0.388	Prob(JB):	0.129
Kurtosis:	3.266	Cond. No.	17.0

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

== 결측치 평균 대체 후 회귀 결과 ==

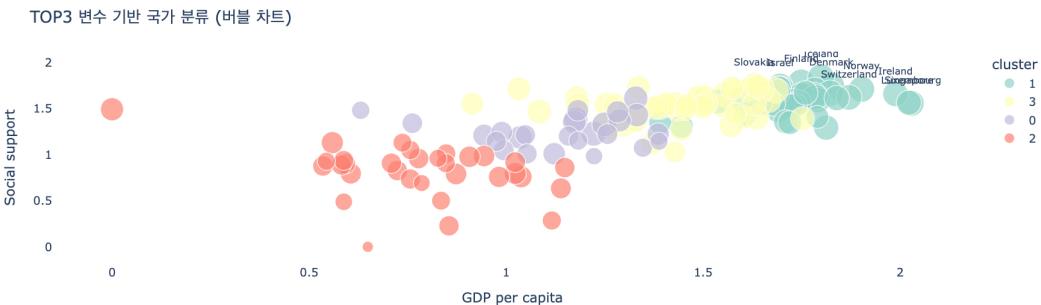
OLS Regression Results

Dep. Variable:	Ladder score	R-squared:	0.756
Model:	OLS	Adj. R-squared:	0.751
Method:	Least Squares	F-statistic:	147.4
Date:	Wed, 26 Nov 2025	Prob (F-statistic):	1.44e-43
Time:	23:32:20	Log-Likelihood:	-126.10
No. Observations:	147	AIC:	260.2
Df Residuals:	143	BIC:	272.2
Df Model:	3		
Covariance Type:	nonrobust		
coef	std err	t	P> t
const	1.7573	0.194	9.063
Social support	1.6978	0.196	8.658
GDP per capita	0.6399	0.210	3.042
Life expectancy	1.2932	0.333	3.885
Omnibus:	4.113	Durbin-Watson:	1.548
Prob(Omnibus):	0.128	Jarque-Bera (JB):	3.670
Skew:	-0.374	Prob(JB):	0.160
Kurtosis:	3.202	Cond. No.	17.0

설명: 다중회귀분석을 통해 Top3 변수의 영향력을 정량적으로 비교한다.

R^2 약 0.75 → 세 변수가 행복지수의 약 75%를 설명하고 있으며, 세 변수 모두 통계적으로 유의미하며 ($P < .01$) 사회적 지지의 회귀계수가 가장 높아 가장 강력한 결정 요인임을 알 수 있다.

5.5 Top3 변수 기반 버블 차트(국가별 균집 분석)



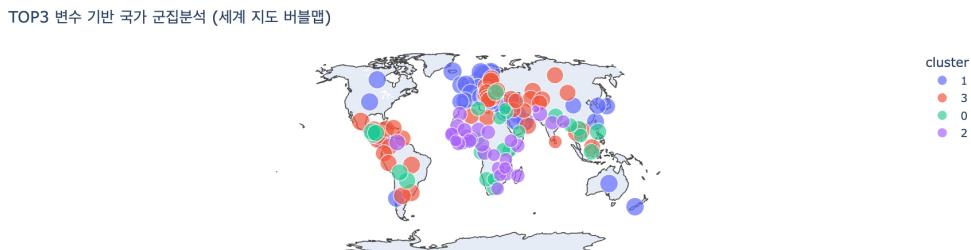
버블차트의 축 의미

축	의미
x축 (GDP per capita)	국민 1인당 경제 수준
y축 (Social support)	사회적 지지(가족/친구/사회 안전망 수준)
버블 크기	행복점수(Ladder score) – 원이 클수록 행복 수준 높음
색깔	K-means 클러스터링으로 구분된 국가 그룹

설명

- 1) 버블이 오른쪽 위에 있을수록, 행복 점수가 높다
오른쪽(높은 GDP) & 위쪽(높은 사회적 지지)에 있는 국가들의 버블이 크고 색깔도 같은 그룹으로 묶이는 경향을 보인다.
- 2) 왼쪽 아래에 있는 국가들은 상대적으로 행복 점수가 낮다
GDP도 낮고 사회적 지원 수준도 낮은 경우, 버블 크기도 전체적으로 작으며, 대표적으로 개발도상국들이 여기에 포함될 가능성이 크다.
- 3) 버블 크기: 행복지수(Ladder score)를 의미하며, 클러스터별 차이를 직관적으로 확인 가능하며, 이를 통해 국가별 행복 수준의 구조적 차이를 시각적으로 비교할 수 있다.

5.6 세계 지도 기반 균집 분석(버블맵)



군집별 특징 (cluster 1> cluster 3> cluster 0 > cluster 2)

(cluster)	특징 요약	버블 크기(행복 수준)	분포 지역
(cluster 1)	GDP 높음, 사회적 지지 & 기대수명 높음	큰 버블 → 높은 행복수준	북유럽, 서유럽, 캐나다, 호주, 뉴질랜드 등

(cluster 3)	경제력 낮거나 중간, 건강/지지 불균형	중간 정도 버블	남미·동유럽·중동 일부
(cluster 0)	사회적 지지와 건강은 중간, GDP 낮음	작은~중간 버블	아프리카·남아시아 중심
(cluster 2)	세 변수 모두 낮은 국가	작은 버블 → 낮은 행복	아프리카 중심, 일부 저개발국

설명

세계 지도 위에 각 국가의 클러스터 결과를 표시해 지리적·지역별 패턴을 분석할 수 있다.

높은 행복도 클러스터(Cluster 1): 미국, 유럽, 호주, 한국 등 경제·건강·사회적 지지가 모두 높은 국가들이 포함된다.

중간 수준 클러스터(Cluster 3, 0): 남미, 동유럽, 중동 일부 국가들이 포함된다.

낮은 행복도 클러스터(Cluster 2): 아프리카 및 취약국(개발도상국)이 포함된다.

세계 지도 기반 시각화를 통해 국가군집이 지역별로 어떻게 분포되는지 쉽게 파악할 수 있다.

5.7 군집별 프로필 레이더 차트

군집별 프로필 레이더 차트 (TOP3 변수 평균 비교)



◆ Cluster 0 국가 목록 (26개)
El Salvador, Guatemala, Paraguay, Philippines, Honduras, Bolivia, Republic of Moldova, Indonesia, Nepal, Lao PDR, South Africa, Gabon, Namibia, Nigeria, Tunisia, Kenya, Uganda, Cambodia, Ghana, Myanmar, Jordan, Egypt, Eswatini, Yemen, Botswana, Lebanon

◆ Cluster 1 국가 목록 (41개)
Finland, Denmark, Iceland, Sweden, Netherlands, Norway, Israel, Luxembourg, Australia, New Zealand, Switzerland, Belgium, Ireland, Austria, Canada, Slovenia, Czechia, United Arab Emirates, Germany, United Kingdom, United States, Taiwan Province of China, Kosovo, Kuwait, Saudi Arabia, France, Singapore, Spain, Italy, Panama, Chile, Malta, Japan, Republic of Korea, Portugal, Cyprus, China, Croatia, Greece, Armenia, Hong Kong SAR of China

◆ Cluster 2 국가 목록 (32개)
Venezuela, Mozambique, Côte d'Ivoire, Congo, Guinea, Cameroon, Senegal, Pakistan, Niger, Morocco, Mauritania, Gambia, India, Chad, Burkina Faso, Benin, Somalia, Mali, Togo, Liberia, Madagascar, Zambia, Ethiopia, Bangladesh, Tanzania, Lesotho, Comoros, DR Congo, Zimbabwe, Malawi, Sierra Leone, Afghanistan

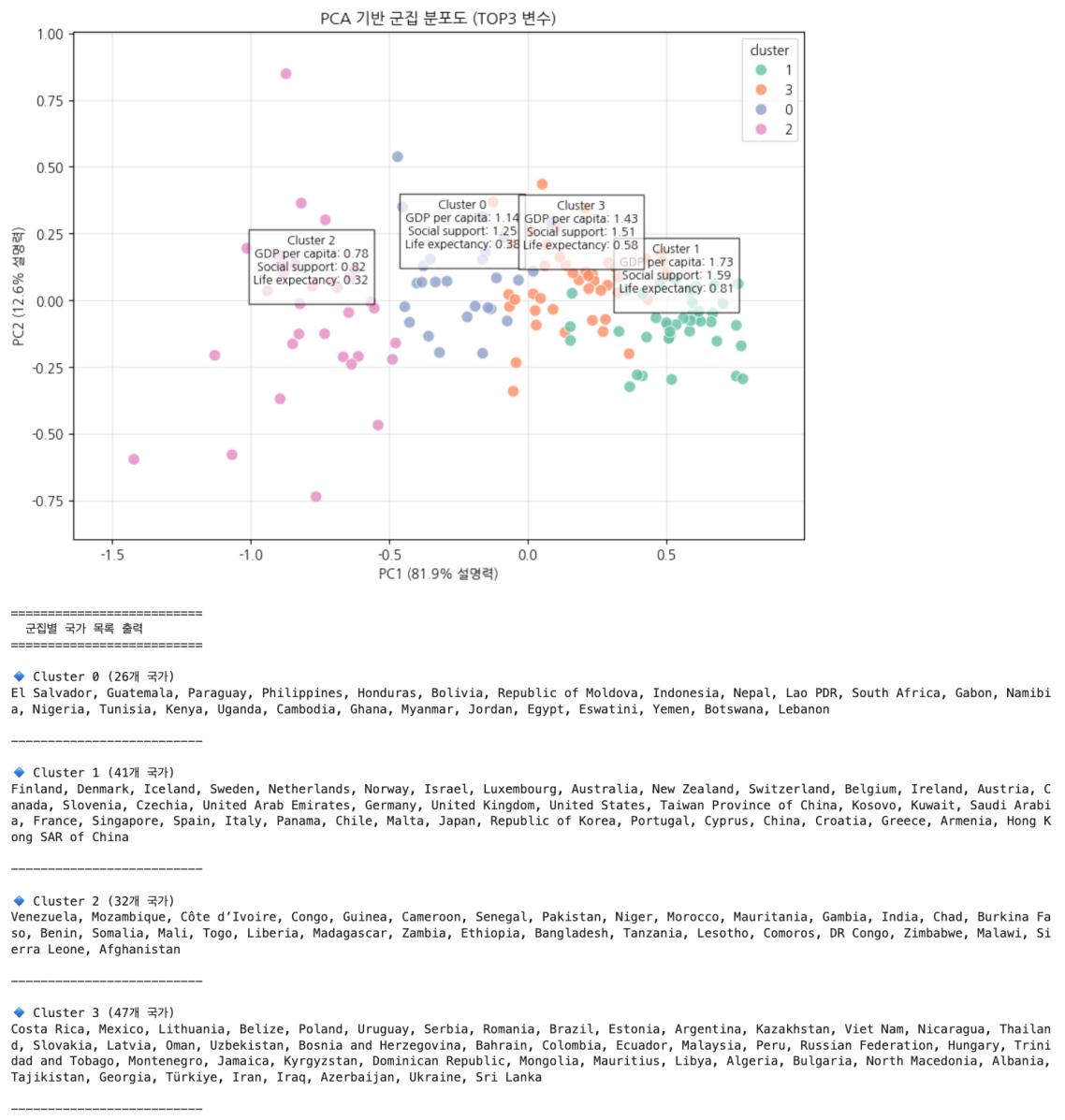
◆ Cluster 3 국가 목록 (47개)
Costa Rica, Mexico, Lithuania, Belize, Poland, Uruguay, Serbia, Romania, Brazil, Estonia, Argentina, Kazakhstan, Viet Nam, Nicaragua, Thailand, Slovakia, Latvia, Oman, Uzbekistan, Bosnia and Herzegovina, Bahrain, Colombia, Ecuador, Malaysia, Peru, Russian Federation, Hungary, Trinidad and Tobago, Montenegro, Jamaica, Kyrgyzstan, Dominican Republic, Mongolia, Mauritius, Libya, Algeria, Bulgaria, North Macedonia, Albania, Tajikistan, Georgia, Türkiye, Iran, Iraq, Azerbaijan, Ukraine, Sri Lanka

각 클러스터의 평균 값을 GDP·사회적 지지·기대수명 3개 축으로 나타낸 레이더 차트이다.

군집(cluster)	분석
Cluster 1	세 지표 모두 가장 높은 값 → 행복 수준 가장 높음
Cluster 3	중간 수준의 균형형 국가
Cluster 0	건강·사회적 지지 약하며 GDP 낮음
Cluster 2	세 지표 모두 낮은 취약군

삼각형의 크기와 형태를 통해 각 클러스터의 강점과 약점을 한눈에 비교할 수 있다.

5.8 PCA 기반 군집 분포도 (TOP3 변수)



설명

PC1 (가로축) – 81.9% 설명력

GDP, 사회적 지지, 기대수명의 종합적 수준의 차이를 가장 잘 설명하는 축이다.

오른쪽으로 갈수록 세 변수(GDP, 사회적 지지, 기대수명) 수준이 높고, 왼쪽으로 갈수록 낮다.

PC2 (세로축) – 10.2% 설명력

상대적인 차이 정도를 보조적으로 표현하는 축, 오른쪽으로 모여 있으면 행복 기반 자원이 풍부한 국가들이고, 왼쪽에 몰리면 그렇지 않은 국가들이다.

5.9 최종 요약

본 프로그램의 결과 화면은 데이터 확인 → 상관관계 분석 → 영향 요인 분석 → 회귀 → 군집 분석 → PCA 시각화까지 전 과정을 시각적으로 확인할 수 있도록 구성되었다.

이를 통해 세계 행복지수의 주요 영향 요인과 국가별 행복 구조의 차이를 체계적으로 이해할 수 있도록 설계되었다.

6. 최종 결론

이 프로그램의 핵심 결론은 다음과 같다.

- 1) 행복을 결정하는 핵심 요인은 경제력, 사회적지지, 기대수명이며, 이 세 변수는 서로 높은 상관관계를 형성하며 국가의 행복 수준을 설명하는 구조적 기반을 이룬다.
- 2) 세 요소 중에서도 사회적 지지가 가장 강력한 영향 요인으로 나타났으며, 이는 행복이 단순한 경제력만으로 설명되지 않는다는 점을 보여준다.
- 3) 군집분석 결과, 전 세계 국가는 세 변수의 수준에 따라 4가지 행복 구조 유형으로 구분되었다.
- 4) PCA 시각화는 군집분석 결과와 일관된 패턴을 보여주며, 국가 간 행복 기반 자원의 차이를 2차원 공간에서 구조적으로 확인할 수 있었다.

6-1. 정책적 시사점

1) 경제적 성장만으로는 행복 향상이 제한적

경제 지표 외적 요소를 함께 강화하는 정책이 필요하다.

2) 사회적지지, 공동체 신뢰 강화가 핵심

다음과 같은 방법이 행복도 향상에 직접적으로 기여할 수 있다.

- 가족, 지역사회 안전망 강화
- 사회적 고립 방지
- 신뢰 기반의 공동체 문화 조성

3) 기대수명을 높이기 위한 보건, 복지 인프라 확충 필요

- 예방 중심 보건정책
- 고령사회 대비 건강 관리
- 의료 접근성 개선

4) 선진국, 북유럽 모델의 ‘균형적 구조’ 참고

경제력·복지·사회 신뢰·자유가 균형을 이루는 국가일수록 높은 행복 수준을 유지한다.
따라서 행복 요인 간 불균형을 점검하고 개선할 필요가 있다.

6-2. 한계 및 추가 분석 방향

1) ‘세계 행복 지수 데이터’ 변수 구성의 한계

‘세계 행복 지수 데이터’는 소득, 사회적 지지, 기대수명, 자유, 관대함, 부패 인식 등 여러 요인을 포함하고 있으나, 행복에 중요한 사회·심리적 요소는 직접적으로 포함되어 있지 않다.

- 정신건강 관련 지표(우울, 불안, 유병률, 자살률), 정치적 안정성, 교육 수준 등

따라서 행복을 설명하는 중요한 일부 요인만 반영하고 있을 뿐 행복의 전체 구조를 포괄한다고 설명하기에는 한계가 있다.

--> 추가 분석 방향

세계 행복 지수 데이터 외에 세계 은행, OECD, WHO 등에서 제공하는 여러 지표들을 결합한 다변량 분석을 수행한다면 행복에 영향을 미치는 구조적 요인을 보다 입체적으로 설명할 수 있을 것이다.

2) 시계열 분석의 부재

본 프로그램은 2025년 단일 연도 데이터를 중심으로 분석에 사용하였다.
이 때문에 행복 수준이 시간이 지나면서 어떻게 변해왔는지, 특정 국가에서 행복 지수가 상승 및 하락한 장기적 추세는 어떠한지 그 패턴을 반영한 해석이 어렵다는 한계가 있다.

--> 추가 분석 방향

다년도의 데이터를 종합하여 시계열 분석을 수행한다면, 장기적인 행복의 추세, 정책 변화나 사회적 사건 이후 행복 지수의 변화 등을 심층적으로 분석할 수 있을 것이다.

3) 문화, 지역적 특성의 반영 부족

이번 군집 분석은 주로 GDP, 사회적지지, 기대수명이라는 정량적 변수 중심으로 국가를 분류하였다. 하지만 실제 행복에는 다음과 같은 문화, 사회 구조적인 특성도 크게 작용할 가능성이 있다.

- 개인주의 vs 집단주의, 가족중심 문화 등 문화권 차이
- 종교적, 가치관적 특성
- 역사, 정치 체제의 차이

이러한 요인들은 숫자로 완전히 표현되기 어렵기 때문에 현재 군집만으로는 왜 특정 문화권이 특정 패턴을 보이는지까지 충분히 설명하기는 어렵다.

--> 추가 분석 방향

동아시아·북유럽·남미·아프리카 등 문화권이나 지역권 단위로 그룹을 나누어 비교 분석하거나, 종교·문화지표를 추가 변수로 활용한다면, 행복 구조에 나타나는 문화·지역적 맥락을 더 풍부하게 해석할 수 있을 것이다.