

1 World Happiness Report

1.1 Exploratory Data Analysis

1.1.1 Data Source & Analysis Objective

본 분석은 World Happiness Report 공개 데이터를 기반으로, 국가별 행복 점수(score)와 관련 지표들의 분포와 관계를 탐색적으로 분석하는 것을 목표로 한다.

데이터는 2015년부터 2024년까지의 국가 단위 관측치로 구성되며, 연도별 표본 수는 약 137–158개 범위에 분포한다. 본 EDA의 목적은 예측이나 인과 추론이 아닌, 데이터 구조와 패턴을 이해하고 이후 분석을 위한 기초 인사이트를 도출하는 데 있다.

1.1.2 Environment & Libraries

본 분석은 Jupyter Notebook 환경에서 수행되었으며, pandas, matplotlib, seaborn, geopandas 등 표준 데이터 분석 및 시각화 라이브러리를 활용하였다. 모든 분석은 재현 가능성을 고려하여 구성되었다.

1.2 Data Loading & Overview

데이터 로딩 이후 컬럼 구조와 연도별 표본 분포를 점검하였으며, 연도 간 표본 수 차이가 분석 해석을 크게 왜곡하지 않는 수준임을 확인하였다.

1.3 Missing Values

결측치는 전체 관측치 중 극히 일부에서만 발견되었다. 결측 비율이 매우 낮고 특정 변수에 국한되지 않은 경우가 많아, 본 EDA에서는 분석 일관성을 유지하기 위해 결측치가 포함된 행을 제거한 뒤 분석을 진행하였다.

1.4 Outlier Exploration

수치형 변수에 대해 box plot을 통해 분포를 확인한 결과, generosity, corruption 변수에서 상대적으로 넓은 분산과 일부 극단값이 관찰되었다. 이러한 값들은 해당 지표의 성격상 국가별 편차를 반영할 가능성이 높아, 본 단계에서는 제거하지 않고 관계 해석 시 참고하는 방식으로 유지하였다.

1.5 Geospatial Visualization Strategy

국가별 행복 지수의 공간적 분포를 직관적으로 파악하기 위해 지도 시각화를 수행하였다. 대표 연도로는 데이터의 시작점이자 이후 시계열 및 코로나 전후 비교의 기준점 역할을 하는 2015년을 선택하였다.

1.6 Geospatial Data Preparation

자연지리 데이터와 행복 데이터는 ISO 코드 기준으로 병합하였다. 분쟁 지역이나 코드 불일치로 인해 일관된 국가 단위 비교가 어려운 경우는 지도 시각화에서 제외하여 해석의 혼선을

줄였다.

1.7 Geospatial Visualization (2015)

1.7.1 World Happiness Rank

행복 순위 지도는 국가 간 상대적 위치를 직관적으로 보여주며, 점수 차이보다 국가 간 격차 구조를 이해하는 데 유용하다. 지역별로 순위가 집중되는 경향이 관찰되며, 이는 지역 단위 사회·경제적 환경의 차이를 시사한다.

1.7.2 World Happiness Score

행복 점수 지도에서는 북유럽 및 북아메리카 일부 국가에서 상대적으로 높은 점수가 관찰된다. 반면 아프리카 및 일부 아시아 국가에서는 낮은 점수가 집중되는 경향을 보인다.

1.7.3 Asia (Asia + Siberia)

아시아 지역을 분리해 살펴보면, 동일 권역 내에서도 행복 점수의 편차가 매우 크다. 이는 단일 지역 평균만으로는 행복 수준을 설명하기 어렵다는 점을 보여준다.

1.8 Visualization Summary (2015)

지도 기반 분석 결과는 다음과 같이 요약된다.

- 지역별로 행복 점수의 뚜렷한 격차가 존재한다.
- 동일 지역 내에서도 국가별 차이가 크게 나타난다.
- 이후 분석에서는 GDP, 사회적 지원, 건강 등 여러 요인을 함께 고려할 필요가 있다.

2 Time Series Analysis (2015–2024)

2.1 Score Distribution: Pre vs Post COVID-19

코로나 이전(2015–2019)과 이후(2020–2024)의 행복 점수 분포를 비교한 결과, 전체 분포 형태는 유사하나 중앙값과 밀도 중심이 소폭 이동한 것이 관찰되었다. 이는 전 세계적 충격 이후에도 행복 점수 분포가 급격히 붕괴되지는 않았음을 시사한다.

2.2 Volatility Analysis

2015–2024 기간 동안 행복 순위 변동성이 큰 국가들을 중심으로 분석하였다. 변동성이 크다는 것은 절대 점수 수준보다는 연도별 상대적 위치 변화가 크다는 의미이며, 이는 국가의 구조적 안정성과 연관될 가능성을 시사한다.

2.3 Rank Trajectory of High-volatility Countries

순위 궤적을 통해 일부 국가들이 기간 내 지속적인 순위 변동을 경험했음을 확인하였다. 이는 특정 연도의 점수만으로 국가의 전반적 행복 수준을 판단하는 데 한계가 있음을 보여준다.

3 Correlation Analysis (Based on Score)

3.1 Overall Correlation Structure (2015)

행복 점수와 주요 변수 간 상관관계를 분석한 결과, `gdp_per_capita`, `social_support`, `healthy_life_expectancy`는 행복 점수와 비교적 높은 양의 상관관계를 보였다. 반면 `generosity`, `corruption`은 상대적으로 약한 상관관계를 나타냈다.

3.2 Heatmap Interpretation

상관관계 히트맵을 통해, 경제·건강·사회적 지원 지표들이 서로 연관된 구조를 가지는 경향이 관찰되었다. 이는 행복 점수가 단일 요인보다는 복합적인 사회·경제적 조건과 연결되어 있음을 시사한다.

3.3 Feature-level Exploration

3.3.1 GDP per capita

GDP가 증가할수록 행복 점수가 상승하는 전반적인 경향이 관찰되지만, 동일한 GDP 수준에서도 국가별 행복 점수의 분산이 존재한다. 이는 GDP만으로는 행복을 완전히 설명하기 어렵다는 점을 보여준다.

3.3.2 Social Support (Family)

사회적 지원 지표는 행복 점수와 양의 관계를 보이나, 사회적 지원이 높음에도 상대적으로 낮은 행복 점수를 보이는 국가가 존재한다. 이는 사회적 지원 외의 요인이 함께 작용함을 시사한다.

3.3.3 Health (Life Expectancy)

기대수명은 행복 점수와 안정적인 양의 관계를 보이지만, 고기대수명 구간에서도 점수 분산이 존재한다. 따라서 기대수명은 중요한 요인 있지만 단독 지표로 해석하기에는 한계가 있다.

3.4 Multi-factor Perspective

GDP, 사회적 지원, 기대수명을 함께 고려할 경우, 여러 요인이 일정 수준 이상 충족될 때 상대적으로 높은 행복 점수가 관찰되는 경향이 있다. 이는 행복이 복합적 요인의 결합 결과임을 시사한다.

4 GDP로 충분히 설명되지 않는 국가들

GDP 단일 요인으로 설명하기 어려운 관측치들을 중심으로 추가 탐색을 수행하였다. 해당 국가들에서는 사회적 지원과 자유 지표가 상대적으로 중요한 역할을 하는 경향이 관찰되었다. 이는 경제적 요인 외 사회적 요인이 행복 형성에 기여할 가능성을 시사한다.

5 Pre vs Post COVID-19 Correlation Comparison

코로나 전후 상관관계를 비교한 결과, GDP는 여전히 행복 점수와 높은 상관관계를 유지하였다. 한편 자유와 부패 인식 지표에서는 소폭의 상관관계 변화가 관찰되었다. 이는 팬데믹 이후 사회적·제도적 요인의 상대적 중요성이 재조정되었을 가능성을 시사한다.

6 Final Summary

본 EDA는 World Happiness 데이터를 기반으로 행복 점수의 공간적 분포, 시간적 변화, 그리고 주요 요인 간 관계를 탐색하였다.

주요 관찰 결과는 다음과 같다.

- 행복 점수는 경제, 건강, 사회적 지원과 복합적으로 연관된다.
- GDP는 중요한 요인 있지만 단독 설명 변수로는 한계가 있다.
- 사회적 요인과 제도적 환경은 상황 변화에 따라 행복과의 관계가 달라질 수 있다.

본 분석은 상관관계 기반 EDA로, 인과관계를 직접적으로 주장하지 않는다. 향후 분석에서는 회귀 분석이나 추가적인 통계적 검증이 필요하다.

A Appendix: Key Visual Evidence for Presentation

본 부록은 발표 시 핵심 주장에 대한 근거를 보조하기 위한 시각화 자료를 포함한다.

A.1 Global Distribution of Happiness (2015)

A.2 Score Distribution Before vs After COVID-19

A.3 Correlation Heatmap with Happiness Score (2015)

A.4 GDP per Capita vs Happiness Score

A.5 Rank Volatility of Countries (2015–2024)

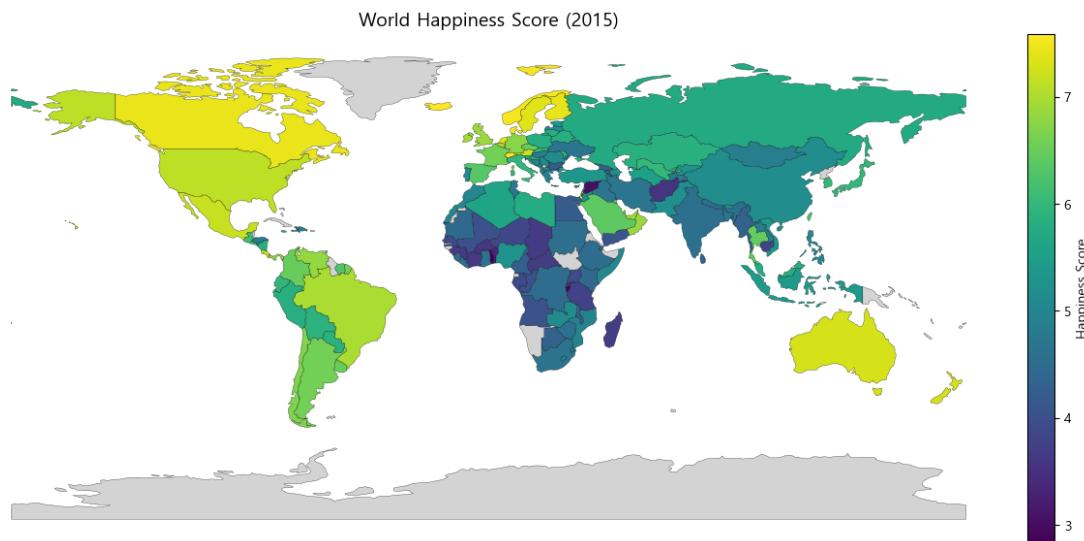


Figure 1: Global distribution of happiness score (2015)

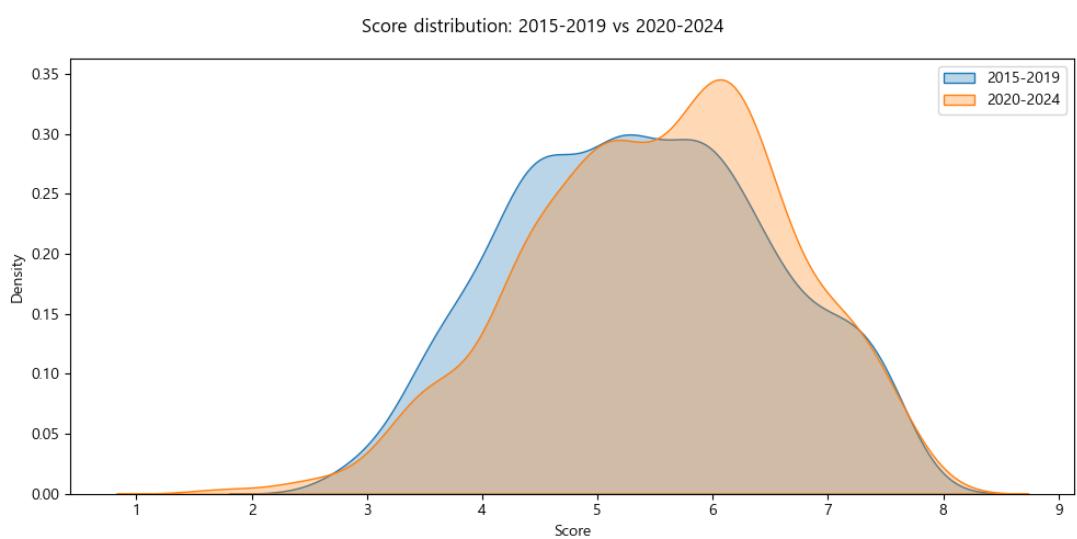


Figure 2: Happiness score distribution: Pre vs Post COVID-19

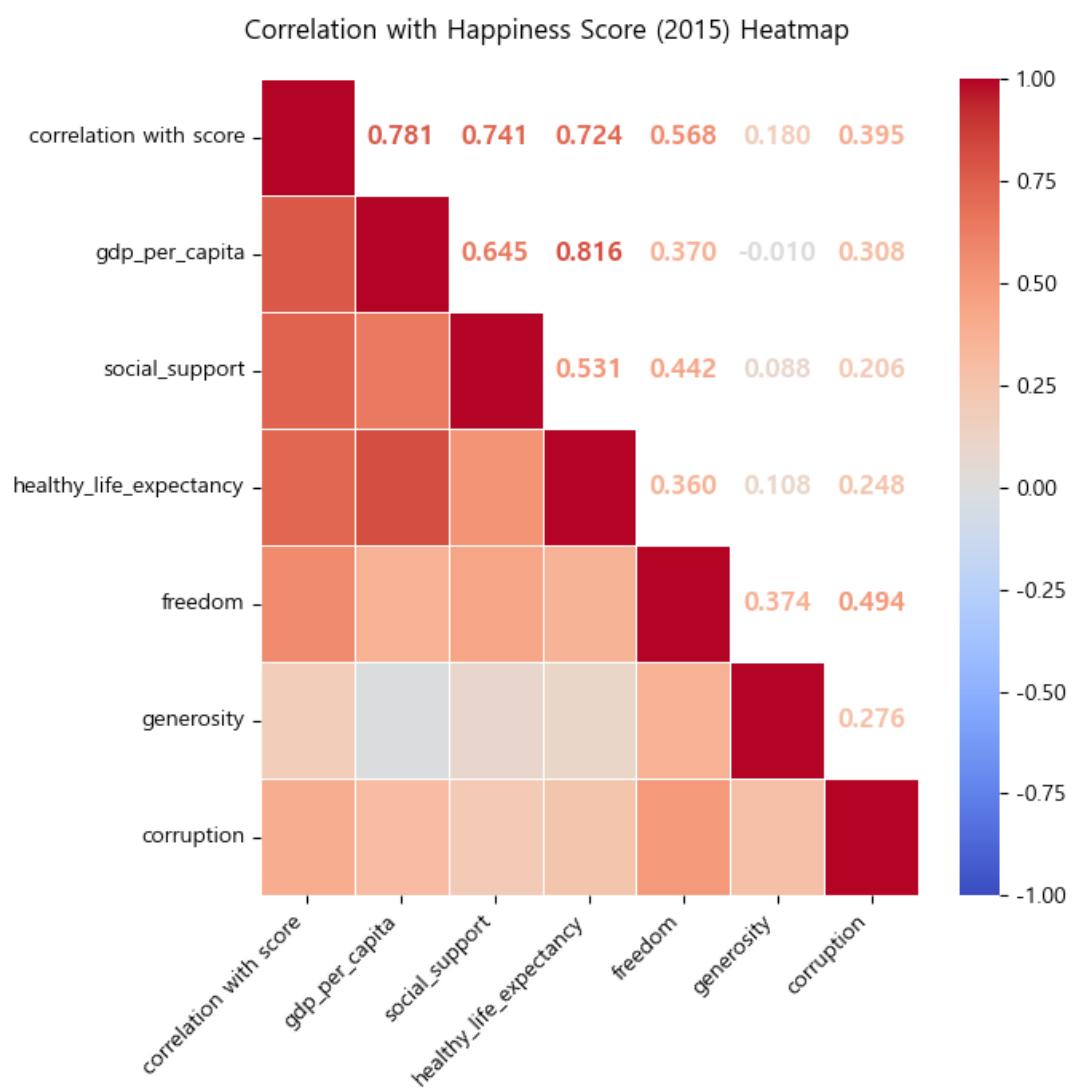


Figure 3: Correlation heatmap with happiness score (2015)

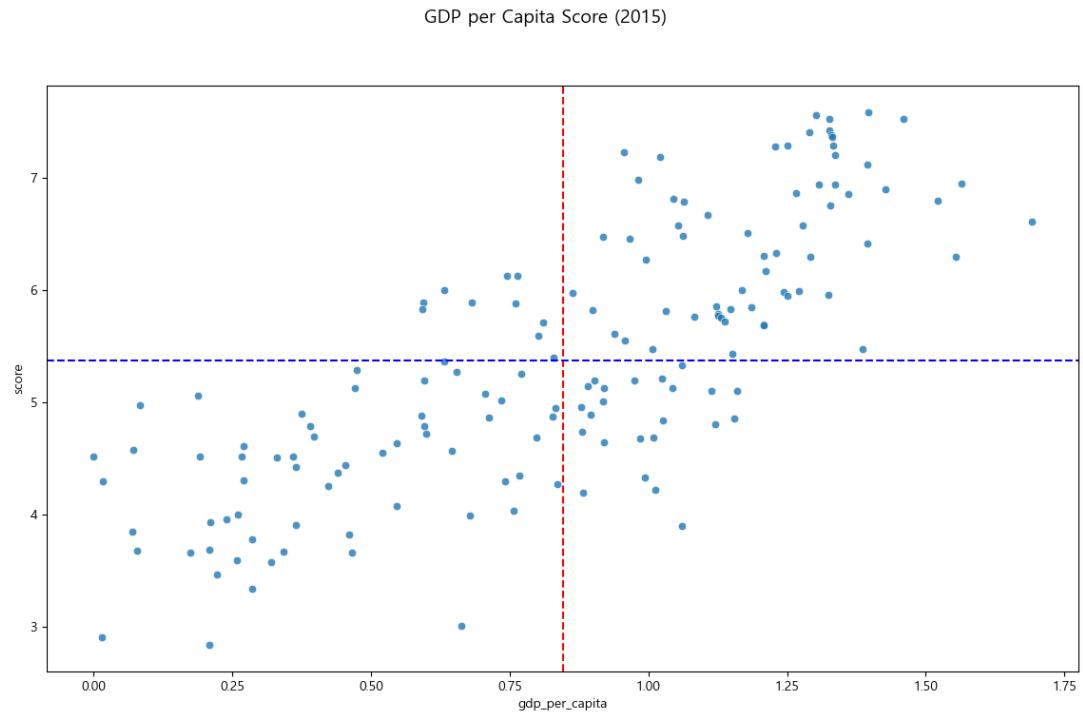


Figure 4: GDP per capita vs happiness score (2015)

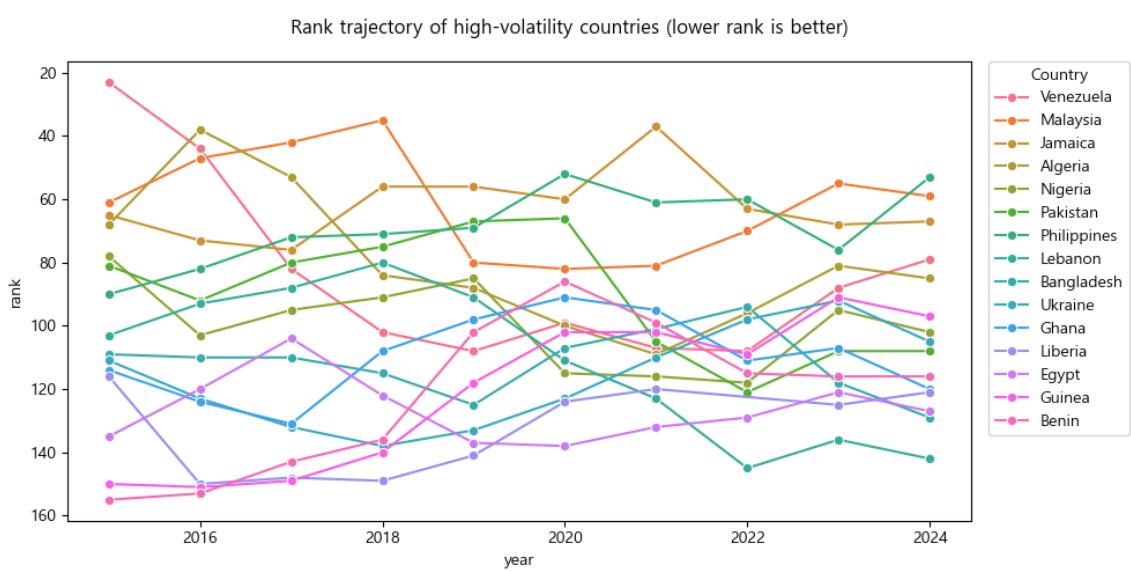


Figure 5: Rank trajectory of high-volatility countries (2015–2024)