

Term Project 결과 보고서

Team7 2019112486 고혜영

2019112463 심미경

1. Dataset

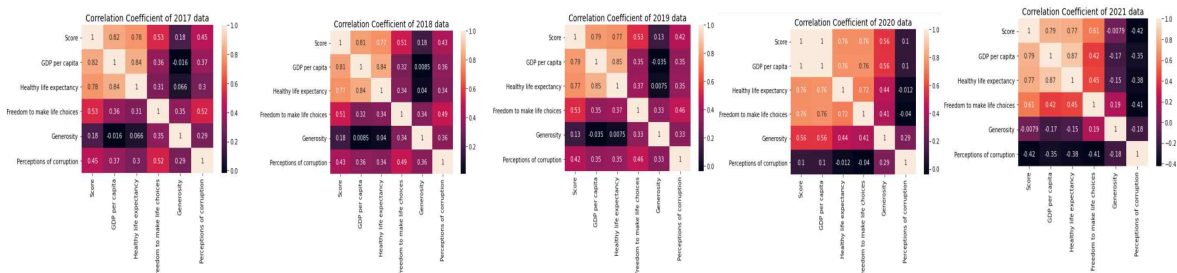
- World Happiness Report 2015-2021 중 2017~2021년도 csv파일 선정
- 각 csv파일 형태: 144행 * 8열
- Dataset 변수(Column명)설명 : readme파일
- 출처: [World Happiness Report 2015-2021 | Kaggle](https://www.kaggle.com/stevenhalpern/world-happiness-report)

2. 주제

- Happiness Score에 영향을 주는 여러 변수 간 관계 분석을 통한 Happiness Score에 가장 영향력이 큰 요소 도출 및 연도별 변화 분석

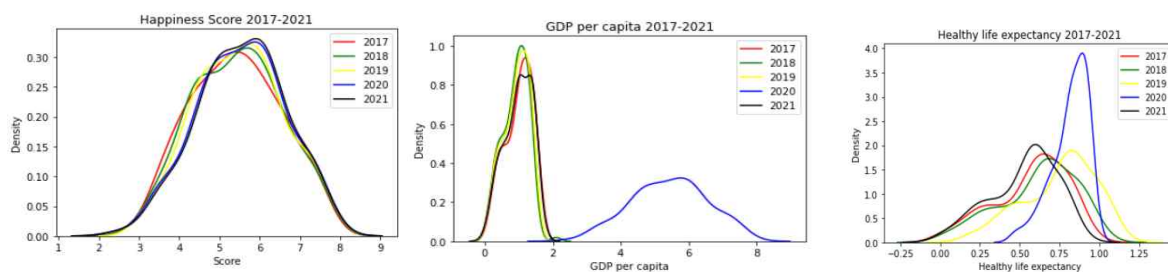
3. 결과 도출 과정 및 해석

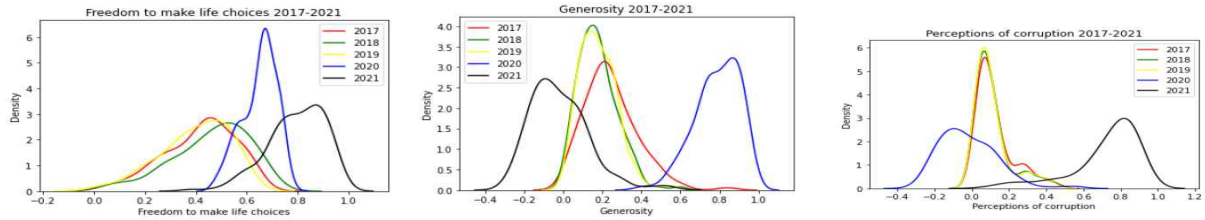
(1) 연도에 따라 변수별로 어떠한 상관관계를 가지고 있는가?



- df.corr()을 통해 각 연도별로 각 변수 간의 상관관계를 도출 후 seaborn의 heatmap으로 나타냈다.
- 2017~2021년 5년간 연도별 상관관계 수치가 다르지만, Happiness Score에 가장 큰 영향을 준 요소는 GDP per capita이며, 가장 작은 영향을 준 요소는 Generosity이다.

(2) 연도별로 변수들의 분포가 어떻게 변화하는가? (계획서 2번 항목 수정)





- seaborn의 kdeplot()을 이용해 각 연도별로 변수들의 분포를 나타냈다.
- Happiness Score는 연도가 지날수록 조금씩 높아지는 경향을 보이고, GDP per capita는 2020년이 다른 연도에 비해 높은 수치를 갖는 것을 확인할 수 있고, Healthy life expectancy는 2020년까지는 어느정도 증가하다가, 2021년에는 가장 낮아졌다. 이는 코로나가 Healthy life expectancy에 영향을 준 것으로 보인다. Freedom to make life choices는 연도가 지날수록 점점 증가한다. 또한 Generosity는 2020년 큰 수치를 가지다가 2021년에 다시 크게 감소한 것을 볼 수 있다. Perceptions of corruption은 2020년도까지 낮은 수치의 분포를 보이다가 2021년에 다시 커짐을 알 수 있다.

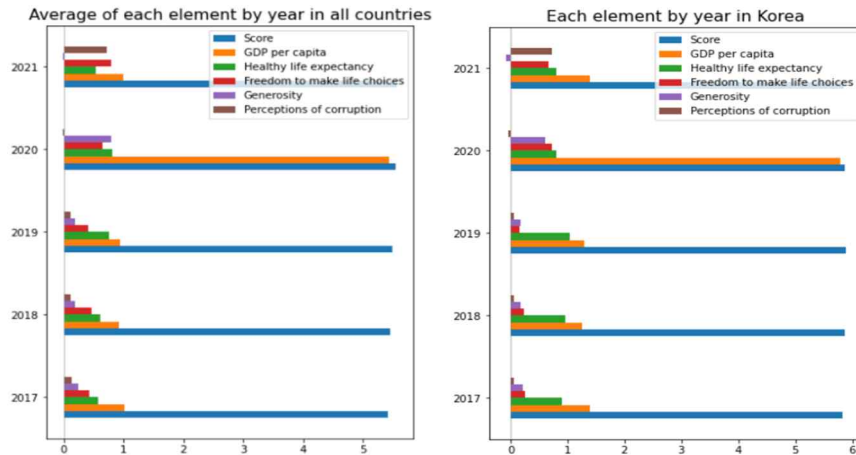
(3) 우리나라의 연도별 각 요소(변수)의 변화는 어떻게 나타나는가?

Korea's Column Values according to the year(2017-2021)

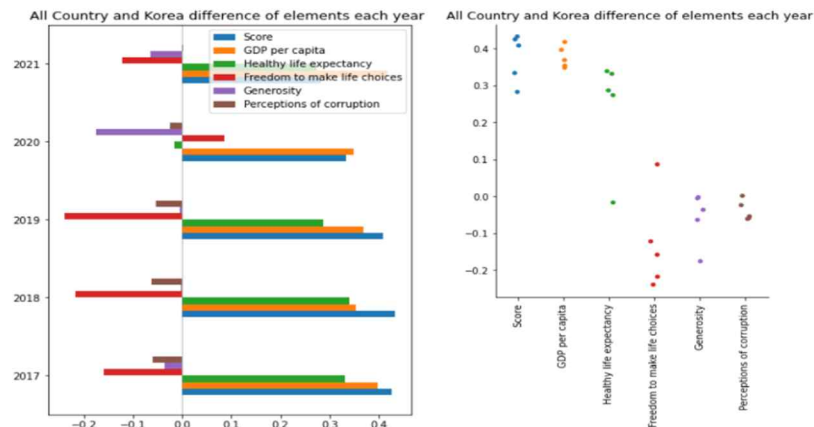


- 매년 달라지는 행복과 관련된 각 요소들의 변화 추세를 line plot을 통해 나타내고, plotly 모듈을 사용하여 그래프에서 marker가 표시된 지점에 마우스 커서를 두었을 때 해당하는 값을 함께 볼 수 있도록 하였다.
- 2020년 'GDP per capita'의 경우 다른 연도에 비해 이상치를 보이며, 이는 2020년 전체 국가에 해당하는 현상으로 해당 연도 데이터의 기록 과정에서 발생가능한 측정 차이로 보인다.

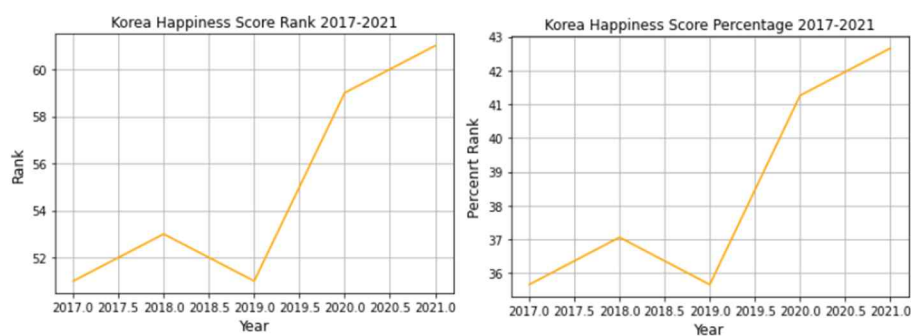
(4) 우리나라가 다른(전체) 국가보다(평균 기준) 높은 요소는 무엇인가?



- 우선, 전체 국가의 연도별 column 값의 평균을 도식화한 barh 그래프와, 한국의 연도별 column 값을 도식화한 barh 그래프를 도출하였다.

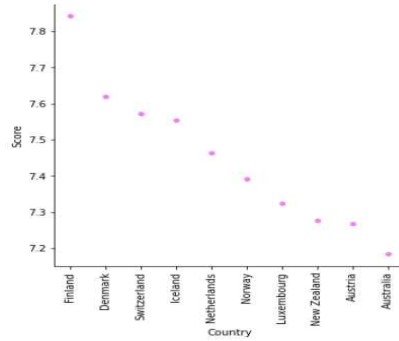


- (5) 우리나라의 Happiness Score를 다른 국가와 비교했을 때 상위 몇 퍼센트에 위치하는가?



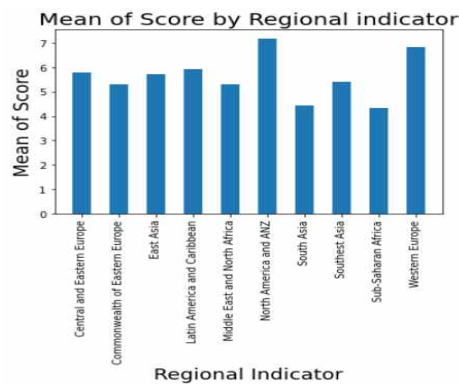
- 우리나라의 연도별 Happiness Score의 순위 변화와 상위 백분위 변화를 line plot으로 나타냈으며 상위 백분위가 순위에 따라 결정되는 값이므로 두 그래프 형태가 동일하다.

- (6) 2021년 OECD국가 중 Happiness Score의 상위 10위는? (계획서 7번 항목 수정)



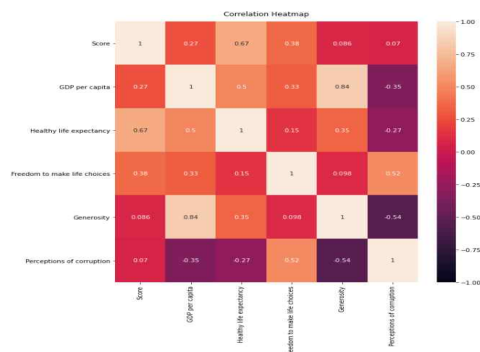
- 5개년의 데이터를 합친 후, OECD국가 목록을 따로 설정하고, 각 국가에 해당하는 열을 Happiness Score가 높은 순서로 정렬한 후 relplot()을 통해 즉, 산점도로 나타냈다.
- OECD국가 중 Happiness Score 상위 10위에 드는 국가는 Finland, Denmark, Switzerland, Iceland, Netherlands, Norway, Luxembourg, New Zealand, Austria, Australia이다.

(7) Happiness Score의 평균이 높은 지역(Regional Indicator) & 낮은 지역(Regional Indicator)



- groupby()를 통해 Regional Indicator별로 Happiness Score의 평균을 도출한 후 x축을 Regional Indicator로 하고, y축을 Happiness Score의 평균으로 하는 barplot으로 나타냈다.
- Happiness Score의 평균이 가장 높은 지역은 North America and ANZ이며, 가장 낮은 지역은 Sub-Saharan Africa인 것을 확인할 수 있다.

(8) Happiness Score에 가장 큰 영향을 주는 요인은? (2017~2021 전체 연도를 고려했을 때)

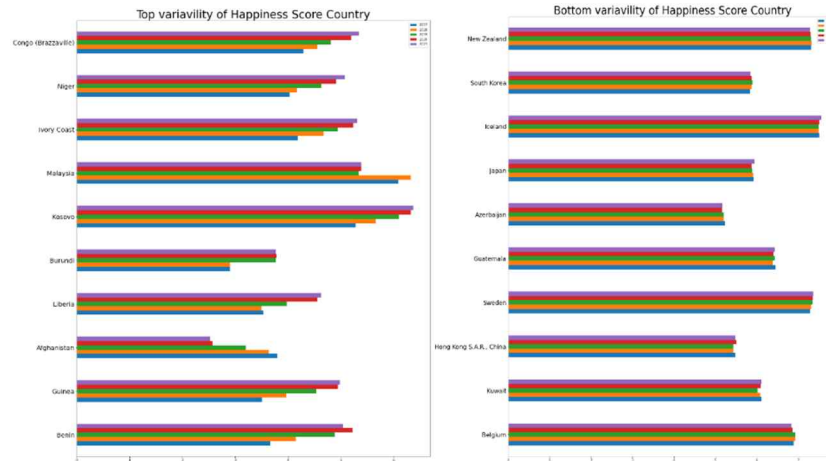


- 2017~2021년의 데이터를 모두 합친 데이터프레임에서 Year컬럼을 없애고 seaborn의

heatmap을 통해 변수들의 상관관계를 나타냈다.

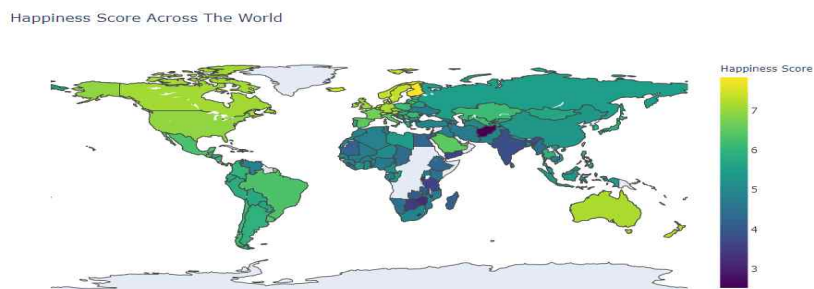
- 2017~2021년의 데이터를 모두 합친 후에는, Happiness Score에 가장 큰 영향을 주는 요인은 Healthy life expectancy인 것으로 나타났다.

(9) Happiness Score가 가장 큰 변동이 있는 상위 국가에는 어떤 것이 있는가?



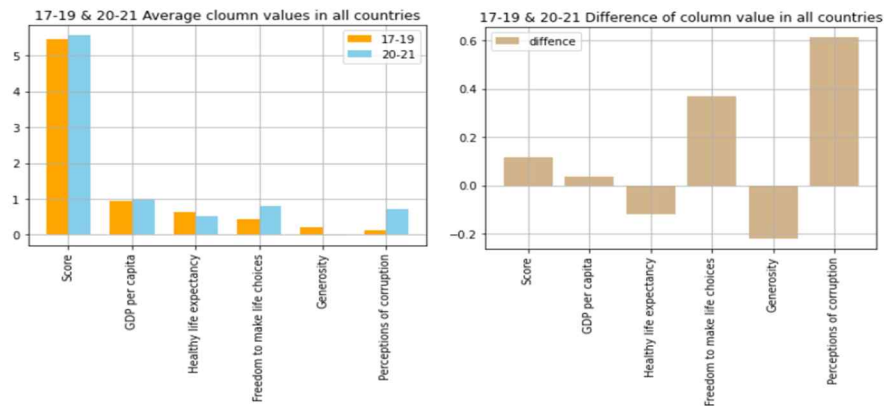
- 각 국가의 연도별 Happiness Score의 변동성 판단 척도를 5년간 Score값의 표준편차로 정의하고, 연도별 Score의 변동성이 가장 큰 상위 10개 국가 및 변동성이 작은 하위 10개 국가에 대한 연도별 Score를 barh 그래프로 도출하였다.
- 변동성이 작은 국가들은 상위 변동 국가들에 비해 연도별 막대그래프의 차이가 매우 작은 것을 확인할 수 있다.

(10) 올해(2021년) 가장 행복한 국가, 즉, Happiness Score가 가장 큰 국가는?

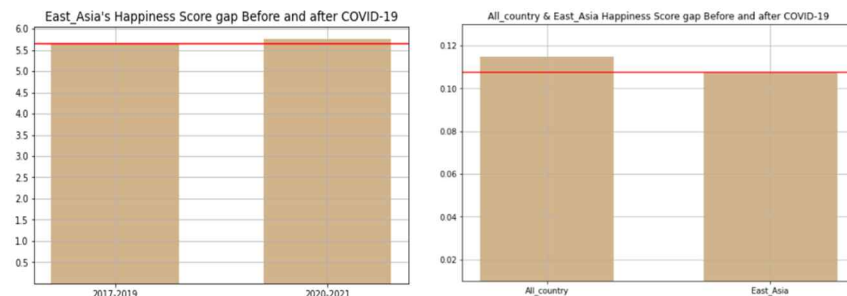


- Country별 Happiness Score값을 plotly의 choropleth map을 통해 나타냈다. choropleth map은 색상이나 패턴을 이용해 특정 통계에 대한 데이터를 시각화한 지도 영역이다.
- 올해 Happiness Score가 가장 높은 국가는 Finland이며, Happiness Score가 7.842이다. 따라서, Finland가 위치한 곳이 노란색으로 나타나는 것을 확인할 수 있다.

(11) 2020년 전후 (코로나로 인해) 전체 국가에서 증가/감소한 요소가 존재하는가?
/Happiness Score에 변화가 있는가?



- 2017~19년과 2020~21년 column별 평균값의 분포(좌)과 그 차이((20~21)-(17~19))(우)에서 Happiness Score가 2020년 기준으로 오히려 증가했음을 알 수 있고, 이를 통해 코로나와 같은 세계적 재해가 행복 지수에 영향을 주지 않았음을 확인할 수 있다.



- 특히 동아시아 국가의 2020년 전후 Happiness Score 변화(좌)를 통해 이 시기에 오히려 Score가 증가했음을 알 수 있다. 또한 전체 국가 및 동아시아 국가의 2020년 전후 Happiness Score 증감 크기(우) 비교를 통해 동아시아 국가에서 Score의 큰 변화가 없었음을, 오히려 전체 국가의 평균에 미치지 못하다는 것을 알 수 있다.

4. 결론

Happiness Score에 영향을 주는 여러 변수 간 관계 분석을 통해서 Happiness Score에 가장 영향력이 큰 요소를 최종적으로 도출할 수 있었으며, 연도별 변화 분석을 통해 여러 인사이트를 도출할 수 있었다.