World Happiness Exploratory Data Analysis ML

Tomás Sousa

February 13, 2021

Contents

Chapter 1

Introduction

In this notebook, we will be analyzing the World Happiness Report from the 2020 dataset as Part of the Exploratory Data Analysis for Machine Learning course by IBM. The idea is to look at worldwide and region-wise trends in happiness score as well as patterns in the importance of the six factors of happiness in determining overall happiness in each country.

With this, we hope to get more hands-on experience in communicating insightful and impactful findings. Throughout this project, we will use the tools and techniques learned in the course.

The datasets used in this analysis are available on Gallup World Poll.

Chapter 2

First Look

In this dataset, we are interested in the Country and Region columns of the dataset in order to group overall happiness scores by country and world region.

Beyond overall happiness score, country, and region, we're also interested in the columns for Economy (GDP per capita), Social Support, Health (Life Expectancy), Freedom, Trust (Absence of Government Corruption), and Generosity. These are the six factors of happiness included in the survey. The values in these columns indicate how much each of the factors contributed to the overall happiness score for each country, and how important they were.

	Country name	Regional indicator	Ladder score	Standard error of ladder score	upperwhisker	lowerwhisker	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption	Ladder score in Dystopia
0	Finland	Western Europe	7.8087	0.031156	7.869766	7.747634	10.639267	0.954330	71.900826	0.949172	-0.059482	0.195445	1.972317
1	Denmark	Western Europe	7.6456	0.033492	7.711245	7.579955	10.774001	0.955991	72.402504	0.951444	0.066202	0.168489	1.972317
2	Switzerland	Western Europe	7.5599	0.035014	7.628528	7.491272	10.979933	0.942847	74.102448	0.921337	0.105911	0.303728	1.972317
3	Iceland	Western Europe	7.5045	0.059616	7.621347	7.387653	10.772559	0.974670	73.000000	0.948892	0.246944	0.711710	1.972317
4	Norway	Western Europe	7.4880	0.034837	7.556281	7.419719	11.087804	0.952487	73.200783	0.955750	0.134533	0.263218	1.972317

Figure 2.1: Dataset first 5 counties

2.1   Data cleaning

After a first analyze was possible to reveal that the dataset was made of 153 countries and 11 features from which we only needed 7 for this exploratory analyze, and did not have NULL values. There were also some features that needed to be resized, in order to give nicest plots, like Life expectancy and GDP per capita.

2.2   Assumptions

After looking at the top 10 countries we could see that 9 were located in Europe's Western Region, indicating that this could be the happiest region on the planet.

Looking at the bottom 10 countries we could see the opposite and the countries with the lowest scores are located in the African Continent. Reporting also low numbers in features such as Freedom, social support, and GDP per capita, compared to the top countries.

Chapter 3
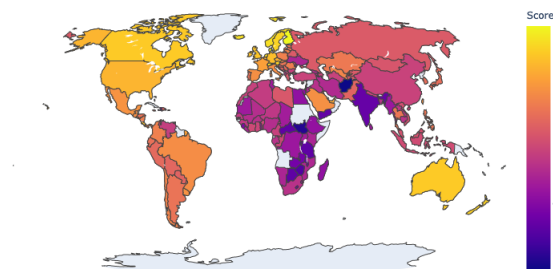
Data Analysis

Happines Arround The Globe -2020

Figure 3.1: Happiness Around The Globe -2020

The lither the colors, the higher the happiness score. Regions in gray do not have happiness data. The happiest regions of the world appear to be in Europe, North, and South America (with the highest area), Australia, and New Zealand. Africa appears to contain the lowest overall happiness scores.

The table of averages and the boxplot both confirm the intuitions we had about the world map for 2020.

In general, Economy and Social support seem to the two most important factors of happiness in these countries. Trust (absence of corruption) and Generosity are the least important.
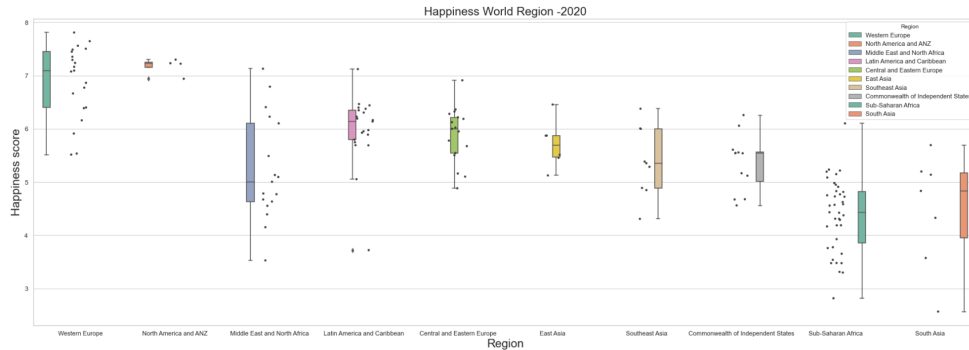

Figure 3.2: Caption

These six factors don't add up to the overall happiness score for each country in the bar plot because they should just be thought of as weights and the 'Dystopian Residual' isn't taken into account.

3.1 Data Distribution

After analyzing the data was possible to verify that the Happiness score report has a normal distribution. On the other side, the majority of the features have logarithmic distributions and polynomial distributions.

3.2 Data Retrieving

3.2.1 Africa

The African continent represents almost 45% of the world's countries and from which there are:

- 5 countries in the Middle East and North Africa that have a happiness score above 6.0

- 2 countries in Commonwealth of the Independent States that have a happiness score above 6.0

- 1 countries in Sub-Saharan Africa that have a happiness score above 6.0

3.2.2 Europe

The European continent represents almost 15% of the world's countries and from which there are:

- 18 countries in Western Europe that have a happiness score above 6.0

- 9 country in Central and Eastern Europe that has a happiness score above 6.0

3.2.3 Asia

The Asian continent represents almost 14.5% of the world's countries and from which there are:

- 2 country in Southeast Asia that has a happiness score above 6.0

- 1 countries in East Asia that have a happiness score above 6.0

- 0 countries in South Asia that have a happiness score above 6.0

3.2.4 America

The American continent represents almost 16% of the world's countries and from which there are:

- 11 countries in Latin America and the Caribbean that have a happiness score above 6.0

- 4 countries in North America and ANZ that have a happiness score above 6.0

One interesting find was the Generosity feature that had a tendency to rank lower in countries with higher GDP and consequently higher Happiness score as it is the case of Finland, the happiest countries on the heath.

3.3 Hypothesis Testing

3.3.1 The correlations and Inferences

in this data set are pretty straightforward as the negative features with greater impact on countries that have positive features rating lower.

For example, countries with higher corruption show lower GDP and lower Life Expectancy, resulting in lower Happiness scores.

One interesting find was the correlation between Life Expectancy with GDP per capita, this could imply that more money means that people leave longer. Of course, this implies other things like higher social support and a possible increase in medical assistance that can not be found in this data set.

3.3.2 Hypothesis

1.
   - Hypothesis A NULL - The Lower the corruption perception the higher the people Happiness report (Ha0)
   - Hypothesis A - The highest the corruption perception the lower the people Happiness report (Ha1)

2.
   - Hypothesis B NULL - The countries with Lower GDP report less generosity (Hb0)
   - Hypothesis B - The countries with higher GDP report less generosity (Hb1)

3.
   - Hypothesis C - Europe is the Saddest region on the hearth (Hc0)
   - Hypothesis C - Europe is the happiest region on the hearth (Hc1)

Looking at this last Hypothesis, let us say that a happy country has a score above 5. So in the entire dataset, we have 100 countries consider to be happy from those 37 are in Europe representing 24% of the world's hippest countries. Nevertheless, the American region also has great scores, with 24 countries reporting happiness.

With this information is not possible to prove the hypothesis once the proves are not strong enough. To do so we need to be biased and reduce the demand of our proves, so if we set the happiness score threshold to 6 then we see some difference from Europe to America so I hold the risk to say that Europe is Happier than America.

Chapter 4

Conclusion

This analysis illustrated that the world's happiest countries are primarily in Western Europe (especially Northern Europe), North America, and Australia New Zealand. It also revealed that the Economy (GDP per capita) is the most important factor in evaluating a country's happiness. Unsurprisingly, the happiest countries and world regions generally tended to be ones with strong and stable economies. The importance of the Economy is also strongly positively correlated with those of Family and Health. This is expected since more economic stability and higher GDP per capita generally encourage stable and comfortable family life as well as increases the availability of proper medical resources and healthcare. These factors then weigh more when determining overall happiness.

We could also hypothesize that these three factors–Economy, Family, and Health–tend to be particularly important because they directly affect individuals living in these countries. Everyone is affected by the state of the economy, especially since it holds direct sway over the availability and security of jobs and the flow of money. Families are the nucleus of home life for most individuals, and Health also affects people on the level of individuals. Consequently, these are very concrete factors and therefore have more influence on the happiness score gauged by individuals.

Sub-Saharan Africa and Southern Asia could definitely use a lift, but overall, the world doesn't seem to be doing too badly.

The next step with this analysis would be to enlarge the data set and see past reports to further predict which features impact the most and what countries can do to elevate their population happiness.