Team Anaconda

02/17/2021

Project 1 Summary and Analyses:

Contents

[1. Happiness Data Report 1](#_Toc64448502)

[2. Unemployment Analysis 1](#_Toc64448503)

[3. Healthcare Analysis 1](#_Toc64448504)

[4. Social Support Analysis 2](#_Toc64448505)

[5. Government Analysis 2](#_Toc64448506)

[6. Weather Analysis 2](#_Toc64448507)

## Happiness Data Report

## Unemployment Analysis

This analysis served to study the correlation between the overall unemployment rate in each country and its happiness score. The analysis was based primarily on two different datasets – one measuring happiness and the other unemployment. Both datasets contained data over the same time frame, 2015 to 2019. The hypothesis being that there is an inverse correlation between unemployment rate and the happiness score.

The datasets came from two different sources and did not contain the same countries. Specifically, the happiness dataset from Kaggle contained information for 141 countries. The unemployment dataset, on the other hand, came from the World Economic Outlook (WEO) database (International Monetary Fund) and included over 200 records. After cleaning and merging these datasets, the final dataset only included 85 records. The merge was conducted using the names of the countries and it is very likely that differences in the ways in which they were spelled in each dataset affected the result.

Nevertheless, analysis showed a moderate correlation between unemployment rates and happiness scores. The r-value of -0.36 demonstrated that there is a negative moderate correlation, in other words, a higher unemployment rate is likely to result in a lower happiness score. The overall correlation calculation is affected by the sample size. Of the 85 records, about six were potential outliers and were excluded from the above calculation. When included, these outliers bump the overall r-value to -0.40.

Another factor that could potentially impact these values is the access of information. Was it easier for happier countries to present information? Were countries with higher employment rates more likely to share information? There is no way to definitively answer this question, however, it is likely that these factors are important. For example, there was no available data on the countries with the lowest happiness scores. Most of these countries had suffered from political turmoil during the timeframe covered in the dataset and it is likely that information was not easily available.

## Healthcare Analysis

This aspect of our project serves to evaluate countries expenditures for Healthcare and their relationship to Happiness scores. To begin this part of the project, data was gathered about I located data on how much each country spends on healthcare, as well as data about populations.

I then divided a countries total spending by population to determine an expenditure per person. I converted this into a csv file. I combined this information with the World Happiness Report Data Index csv file into 1 dataframe. I then cleaned the merged data and renamed any columns to make the dataframe fields more readable. I ran a regression analysis to determine correlation between healthcare spending per person and happiness score.

I found a strong correlation between the two. It made me curious if only highly populated areas spent the most on healthcare so I did an analysis just between population and happiness.

What I found was there was no correlation between happiness and population, so my first finding, spend on health per person, told the real picture.

## Social Support Analysis

## Government Analysis

To find the types of governments on the top and bottom ten countries, I used the Database on Political Institutions (DPI) from the World Bank Data Catalogue to extract the types of political systems with in each country during 2015 to 2017. Merged this information with the average happiness per country taken from the "world Happiness Report"

I was interested in looking at what types of governments were the top ten and bottom ten average happiness countries to further understand if this was a contributing factor to the happiness ranking/ score.

I used the bar plot to have a visual of the 3 types of government used in the in the DPI form the world bank database. Described below:

* “0”: Direct Presidential are Countries with presidents who are elected directly or by an

electoral college

* “1”: Assembly Elected President are Countries where the selection of the effective executive

is done by an elected assembly or by an elected but uncommitted electoral college

* “2”: Parliamentary are Countries in which the legislature elects the chief executive

Findings:

There is a positive relation between the quality of government and the average happiness.

When sorting through the data on the top ten countries ranked by highest average happiness, these countries are all under Parliamentary form of Government. While on the bottom 10 countries 8 out of 10 are Direct Presidential.

## Weather Analysis

To determine the weather impact of a country on their happiness score the happiness data served as the first input. In the jupyter notebook “weather.ipvnb” countries were combined with a list of the nations’ capitals and for each capital, a current weather report was pulled. The question to be answered with this aspect of our project is not what weather influences happiness, but what weather can be expected were we to visit one of the ten on average happiest countries from the World Happiness Report.

The capital locations with their respective countries were processed in preparation to be merged. The city names were then sent out with a request for current weather conditions from an API. The returned data was converted into a new dataframe for further evaluation and exported as a .csv file. Further cleaning of data was completed to prepare for a merge with the happiness data report. Cities lacking weather data (“NaN” in city field) were removed, as were any empty data rows, of which there were none. City weather data was merged back into countries and capitals list, to allow further processing by country instead of by city. Here also, any row that had no value or had an “Nan” fields were removed from the dataset. This dataframe was then again merged with the Average Happiness data file from section 1, which provides average happiness over the four years 2015 – 2019 in one score. This score was used to slice the weather data into a dataframe showing weather for the top 10 happiest countries worldwide. The resulting dataset “happiness\_weather\_top10” was exported as a .csv file for the following visualizations.

Visualizations were prepared to depict any corresponding trend for current weather metrics. It is, of course, winter in most of our regions to temperatures were very low, except for in Oceania, where it is currently Fall.

* **Outline the goals (scope) and purpose of your project**

*Our goal is to understand on a data-driven level the factors contributing to and correlating with World Happiness. We will examine the relationships between several metrics of individual countries or regions and their happiness scores or ranks to determine their influences and effects.*

* **Create a brief summary of your interests and intent:**
  + What kind of data would you like to work with, or what field are you interested in?

*We will utilize the World Happiness Report dataset at Kaggle.com to perform an analysis of different factors contributing to, detracting from or correlating with world happiness.*

* + **What kind of questions could you ask of that data?**

*We will be asking ourselves how a country’s economy, life expectancy, unemployment, weather, type of government, access to healthcare, family size and access to technology interact with that country or region’s happiness score.*

*We will be looking for factors showing the greatest or least amount influence on a country or region’s happiness score.*

*We will be evaluating correlations between the different data sets described above and the world happiness report.*

*Some questions we will be asking:*

1. *What factors play a role in the happiness score of a country?*
2. *What is the correlation between each factor and the happiness score?*
3. *What is the development of a country or region’s happiness over time?*
4. *What is different between the highest ranked countries and the lowest?*
5. *What differences can be found between regions of the world?*
   * **What sources might you find this data?**
     + [*https://www.kaggle.com/unsdsn/world-happiness*](https://www.kaggle.com/unsdsn/world-happiness)
     + *Other data sources as available to answer the above questions.*

Results: TODO each group member summarizes their findings and hypotheses

What did we find for each factor? How did we go about coming to our conclusions?

1. Healthcare Spending
   1. The more governments spend, the happier they are
      1. Explain outliers, process of coming to above hypothesis.
2. Unemployment
3. Government Type
4. Social Support
5. Weather
   1. Weather data is retrieved in real-time, differentiating it from the other factors, which are of a historical nature.