

**TASK**

**Exploratory Data Analysis on the World Happiness Report Data Set**

[](http://www.hyperiondev.com/portal/)

**Introduction**

In this EDA, the main objective was to look at the comparisons of dataset of World Happiness Report. We are comparing two datasets from the years 2015 and 2019. The focal point is around checking the happiness scores and how they are affected by other factors involved. The data used is from the Gallup World Poll. The scores are based on answers to the main life evaluation asked in the poll. The columns following the happiness score estimate the extent to which each of the six factors – economic production, social support, life expectancy, freedom, absence of corruption, and generosity- contribute to making life evaluations higher in each country than they are in Dystopia, a hypothetical country that has values equal to the world’s lowest national averages for each of the six factors. They have no impact on the total score reported for each country, but they do explain why some countries rank higher than others.

**DATA CLEANING**

# SUMMARY OF THE METHODS AND VISUALIZATIONS DONE DURING DATA CLEANING

* In both the 2015.csv and 2019.csv we found out that the data given is clean. We did little things like checking for duplicates across all the rows per column. We looked on the country and region for the 2015 dataset since they’re the only tables that are having non-numeric values. This was to check for inconsistencies. We tried to do the following:
* Convert everything to lowercase
* Remove duplicates
* Remove white spaces
* We didn’t notice inconsistencies in the region/country column(s) and as such didn’t use a module called FuzzyWuzzy for string matching.

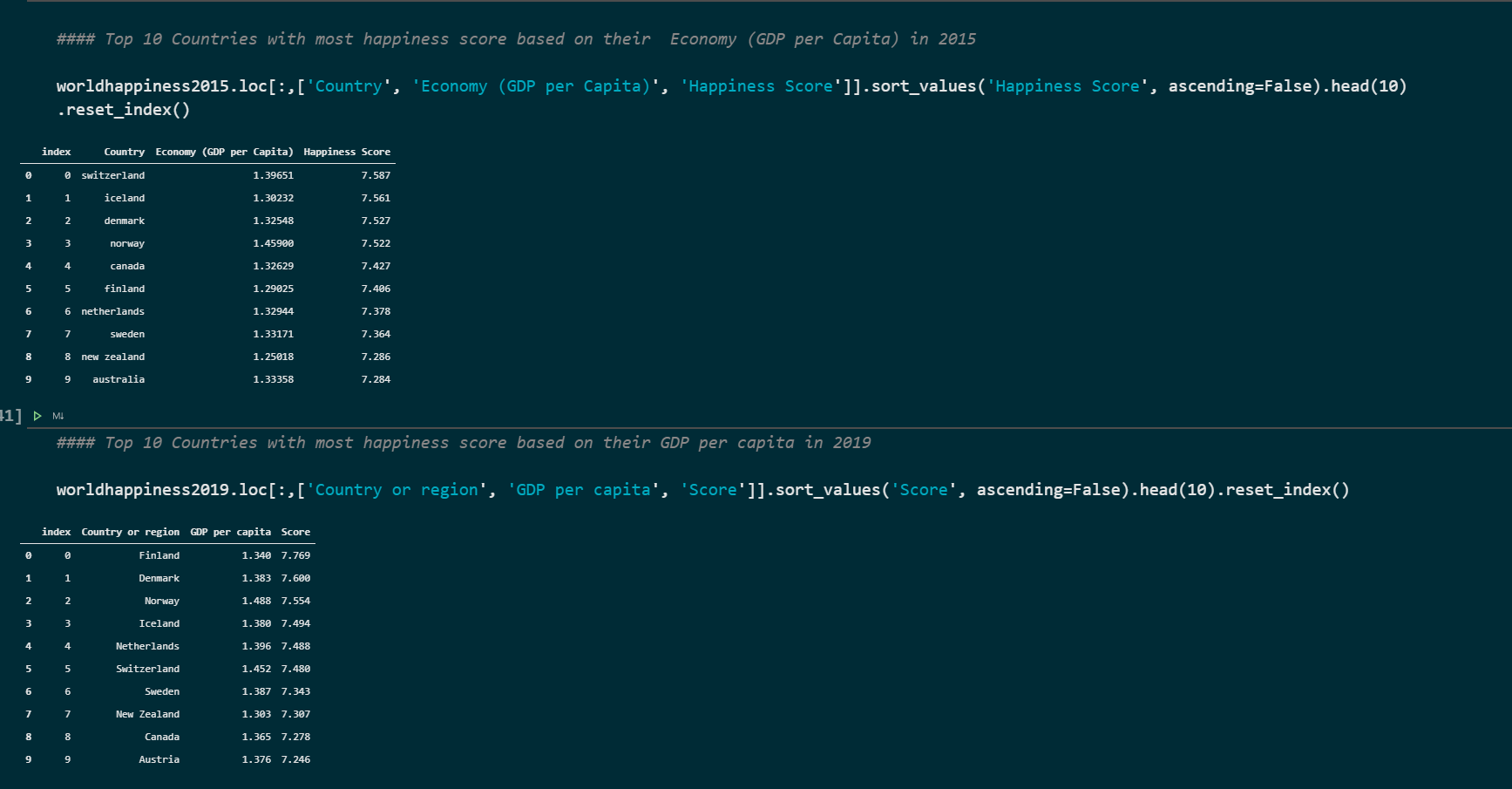
MISSING DATA

# ANY MISSING DATA? HOW DID YOU HANDLE IT

* It was found that there’s no missing values in this dataset.

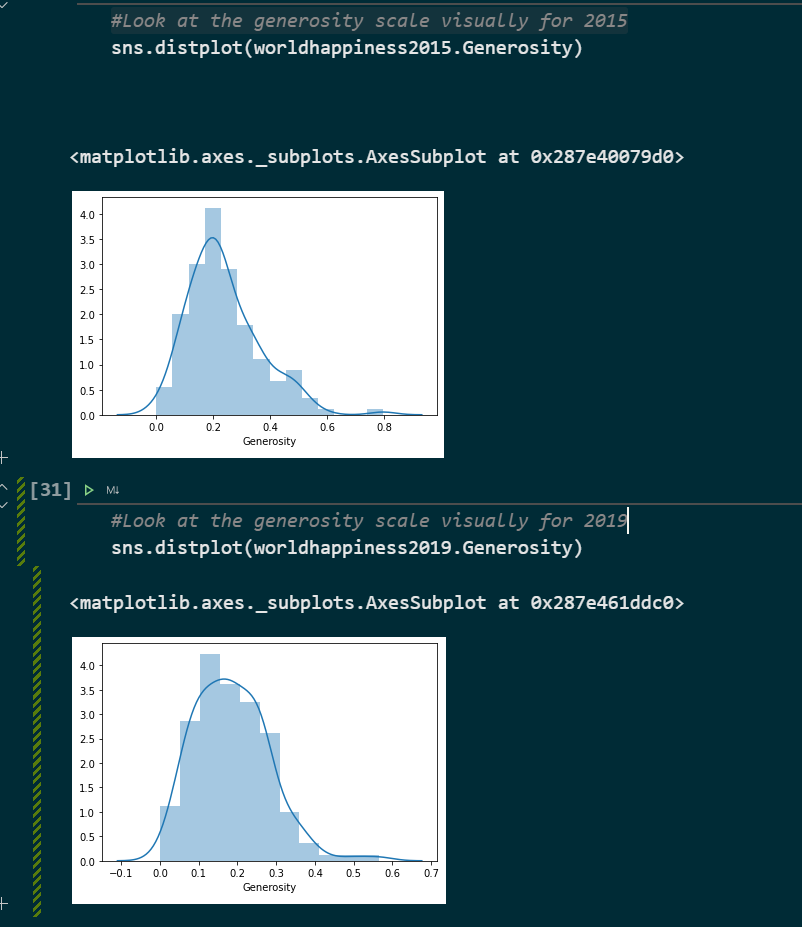
DATA STORIES AND VISUALIZATIONS

1. We compare the top 10 countries with most happiness score based on their economy (GDP per Capita) in 2015



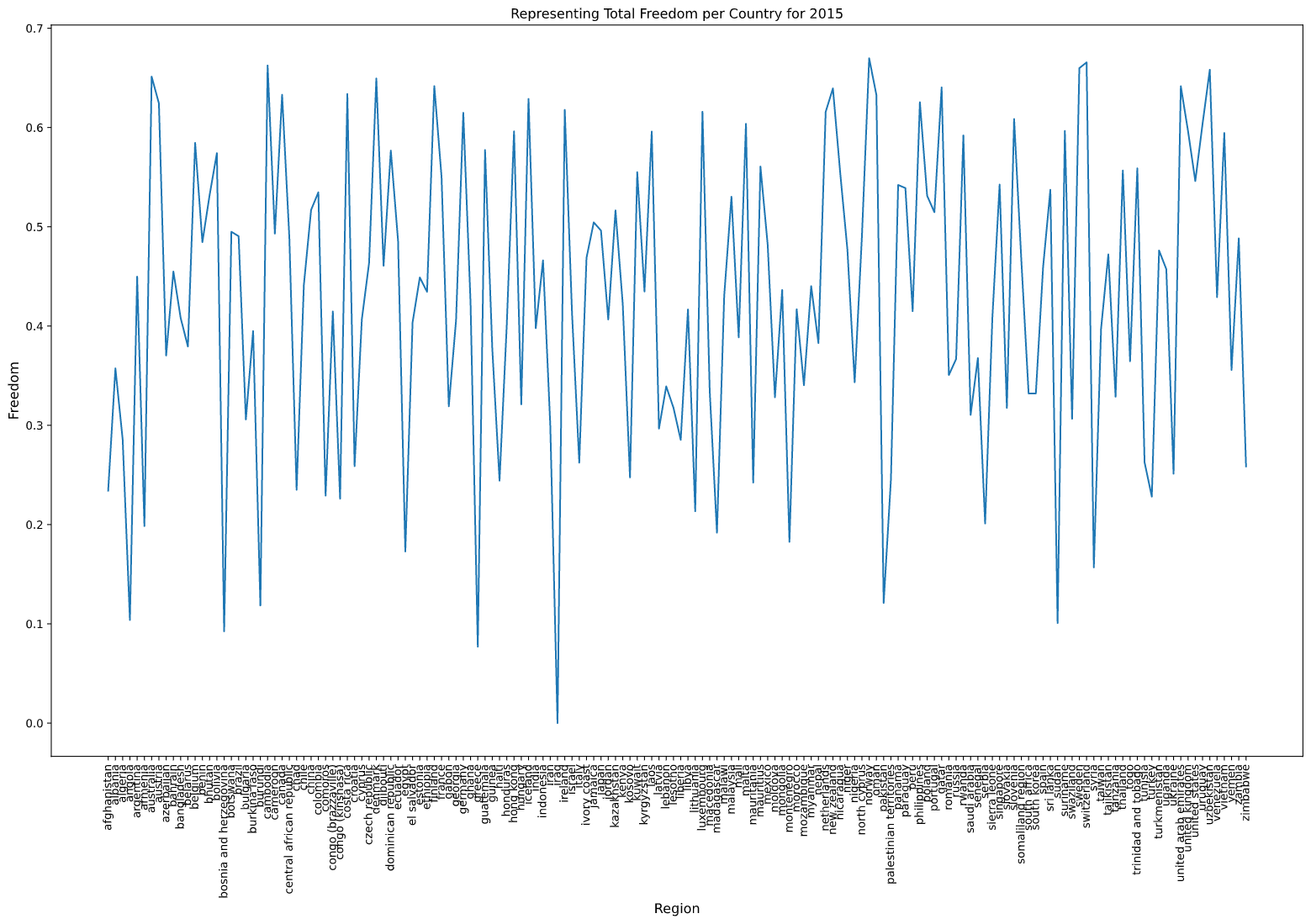
* It is evident from the comparison that Austria is having the least score in both datasets taking position 10. It is having scores of 7.284 and 7.246 respectively in 2015 and 2019.
* Switzerland dropped in both its GDP and Happiness Score in 2019, taking position 6 losing its leading rank from 2015

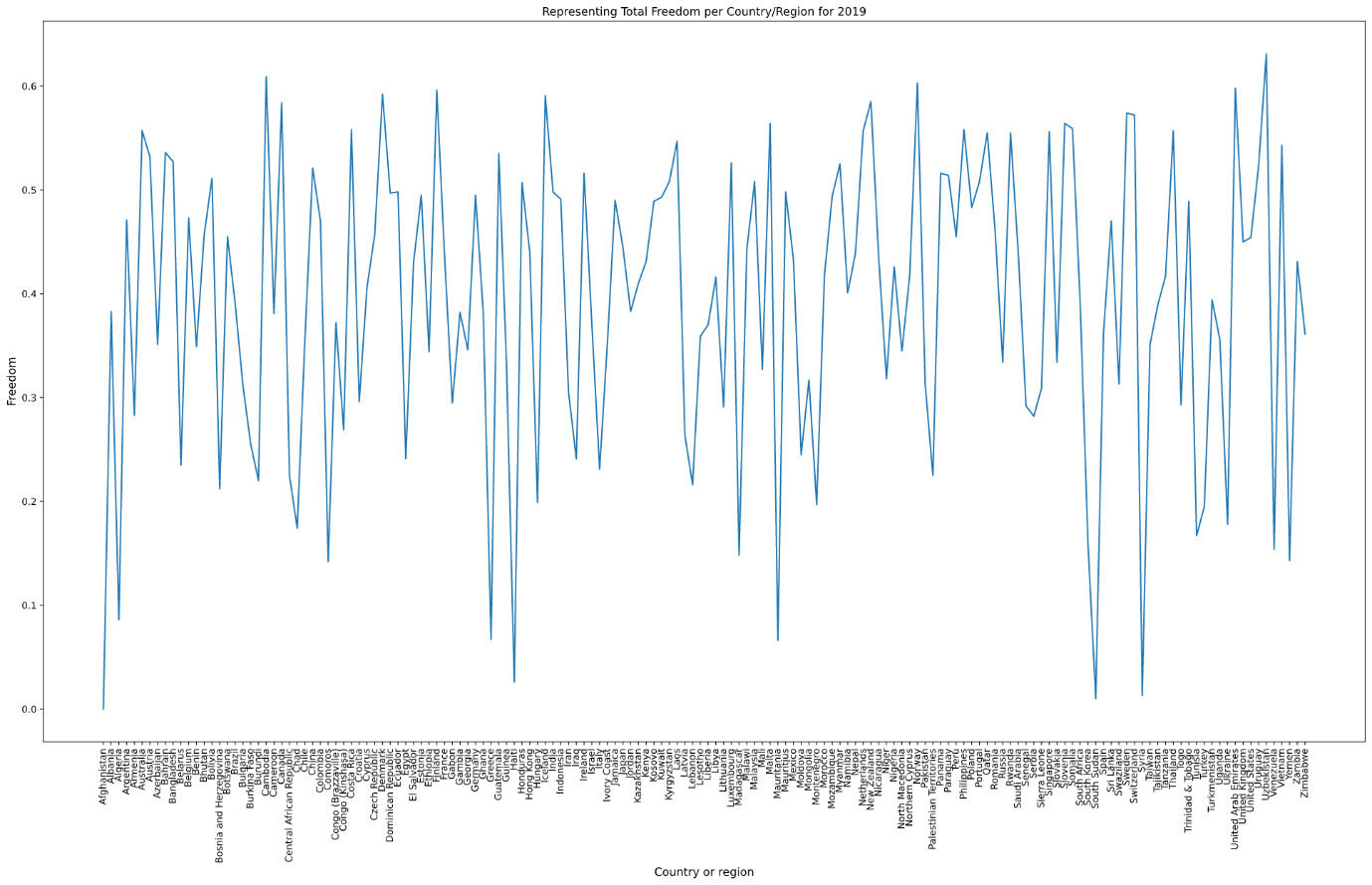
1. We look at generosity scale in both 2015 and 2019

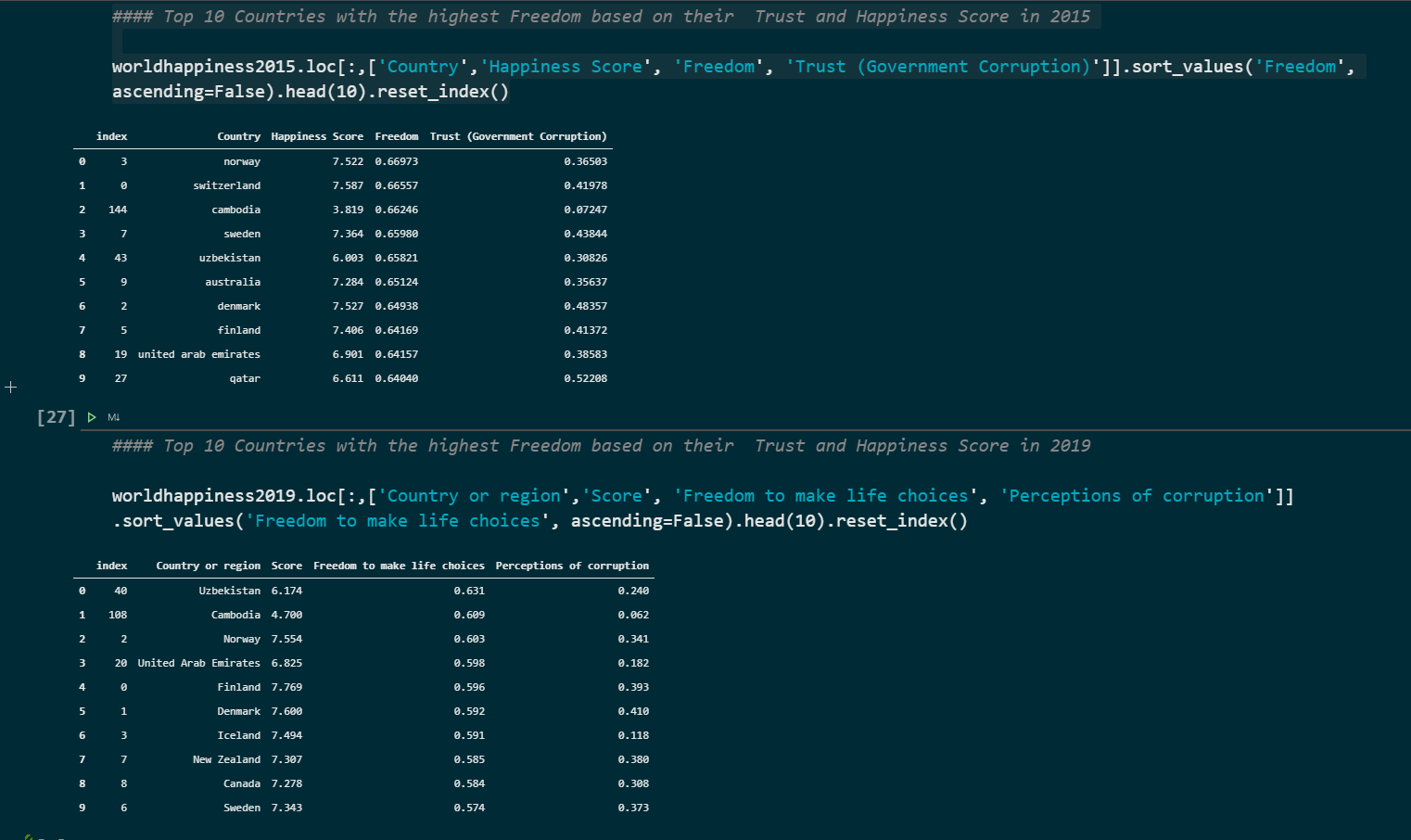


1. Representing total freedom per country/region in 2015 and 2019

This enables us to look at how this freedom affects the happiness score and absence of corruption



****



* From the above data, we can clearly see that having the highest freedom doesn’t necessarily mean that the trust (government corruption) is high. Looking at 2015, Norway is having the highest freedom yet the Trust was amongst the lowest. In real life we know that where there is too much freedom there is corruption and the perception of this trust in most cases is high.

**THIS REPORT WAS WRITTEN BY : LIVHALANI DAU**

