Project Report

# GitHub URL

https://github.com/sdaubers/UCDPA\_saradauber.git

# Abstract

This Data Analytics assignment focuses on the World Happiness Report from March 2022. The World Happiness Report is a publication of the [United Nations](https://en.wikipedia.org/wiki/United_Nations) [Sustainable Development Solutions Network](https://en.wikipedia.org/wiki/Sustainable_Development_Solutions_Network). It contains articles and rankings of [national happiness](https://en.wikipedia.org/wiki/Gross_National_Happiness), based on respondent ratings of their own lives, which the report also correlates with various [(quality of) life](https://en.wikipedia.org/wiki/Quality_of_life) factors. As of March 2022, [Finland](https://en.wikipedia.org/wiki/Finland) had been ranked the 'happiest' country in the world five times in a row. The report primarily uses data from the [Gallup World Poll](https://en.wikipedia.org/wiki/Gallup_(company)#Gallup_World_Poll). (Source [Wikipedia World Happiness Report](https://en.wikipedia.org/wiki/World_Happiness_Report#2022_World_Happiness_Report))

# Introduction

I have chosen to focus on this topic of World Happiness due to the turmoil and unrest which the world and peoples’ lives has been exposed to over the past few years due to serious and once-per-century events, namely Covid and war.

The World Happiness Report examines the Happiness across the globe based on factors such as GDP per capita, Generosity, Healthy Life Expectancy, Social support, Freedom to make life choices and Perception of Corruption.

The objective of my analysis was to examine potential impacts from other factors on the global happiness. I chose to focus on two relevant topics for this analysis, namely Covid19 and Work-Life-Balance. I have used Python 3 as the language for my analysis and Jupyter Notebook(anaconda 3) as the editor.

# Datasets

I have chosen the following datasets for my analysis:

[World Happiness Report up to 2022 | Kaggle](https://www.kaggle.com/datasets/mathurinache/world-happiness-report?select=2022.csv)

* The most up to date information on the World Happiness report from March 2022

[World Happiness Report 2021 | Kaggle](https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2021)

* A reliable source on Kaggle for the World Happiness report from March 2021

<https://api.covid19api.com/summary>

* This API was referenced in the API examples from our lecture notes
* This data contains the most up to date Covid information per country i.e. covid-related deaths and cases

[Countries in the World by Population 2022 | Kaggle](https://www.kaggle.com/datasets/anandhuh/countries-in-the-world-by-population-2022)

* This dataset lists the population per country globally

[Cities with the Best Work-Life Balance 2022 | Kaggle](https://www.kaggle.com/datasets/prasertk/cities-with-the-best-worklife-balance-2022)

* This dataset displays the work-life-balance scores of global cities and their countries. I chose this source as Kaggle is reliable and I believed this information to be interesting to analyse in conjunction with world happines.

# Implementation Process

**Preparation phase**

As the first step of the preparation for the analysis, I imported the required libraries for my analysis and for the visualization of my results. Python libraries are packages or collections of various modules containing functions and methods which contain reusable code in order to reduce the quantity of code you need to write in your own program.

**Datasets for Analysis Part I – World Happiness Report 2022 vs 2021**

I chose the World Happiness Report 2022 and also the 2021 report as the first datasets for my analysis.The 2022 report is the most up to date data which was released in March 2022, the 2021 report in March 2021. The following were the steps which I carried out:

1. Import of 2022 and 2021 dataset files as csv using panda .read() function and viewed the first 5 rows using .head()
2. Perform data cleansing on the 2022 data by checking for NA data i.e. empty data and drop if applicable
3. Next I wanted to show the max and min values per category in the 2022 data. For this I created functions called show\_max and show\_min which return a background colour (green for max, red for min) when the function is called by the style function apply() and applied to the dataset
4. As a next step, I wanted to show the top 10 countries per category included in the Happiness Report 2022:
   * I defined my subplots using matplotlib pyplot, and defined how they should be visualized i.e. 3 rows and 2 columns.
   * I wanted to use the nlargest function to create the subset this function only accepted integers and my data were all objects and containing commas
   * To solve this, I needed to first replace the commas using the str.replace() function and then convert the data to int using astype() function
5. View the Top and Bottom 10 using the sort\_values and the list end range set to 10. I used Seaborn Barplot
6. Next, I analysed how Ireland fares in the report results, overall in the world and also in Europe:
   * Used the .loc function to just look at Ireland and saw that it ranked number 13
   * In order to analyse Western Europe, I needed to merge the data with the 2021 data to retrieve the Regional Indicator
   * Created a ranking for Western Europe.
   * The result was that Ireland was number 9 in Europe but 13 in the world
   * Then I wanted to compare Ireland with the number 1 country in Western Europe and the world: Finland. I used the isin() function on the Country name category to filter for Ireland and Finland. Ireland falls down on the categories ‘Perceptions of corruption’ and ‘Freedom to make life choices’.
7. Finally, I wanted to visualize and compare the top 10 countries from 2022 vs 2021. For this I used seaborn barplots and the nsmallest() function on the category ‘RANK’. This showed that the main movers were: Austria which lost it’s spot in the Top 10 and Israel which joined the ranks in the Top 10 at number 9

**Datasets for Analysis Part 2 – World Happiness Report 2022 & Covid19 API & Population Dataset**

As my second analysis, I was interested in examining the potential impact of Covid19 on the World Happiness Report. For this, I used a Covid19 API which I imported using requests.get() function and the .json() function which then enables the passing of the information back and forth to the API. I assigned this information to a Pandas Dataframe to continue the processing for analysis:

1. Data cleansing: checked for NA values which resulted in no results
   1. Renaming the column Country to Country name in order to merge the data with the World Happiness dataset
   2. Renaming other columns to be more descriptive e.g. ‘TotalConfirmed’ -> ‘Total Covid Cases’
2. Merging of the Covid19 dataset with the World Happiness 2022 report in order to visualize the potential impact involved the following steps:
   1. Took a subset of the Covid19 data containing just the relevant columns for analysis and displayed the Top 10 countries for deaths and cases
   2. Merged the Top 10 subset with the World Happiness Data for 2022, on the Country name and using a left join in order to ensure preference for our Covid dataset
      1. This returned some missing rows/unsuccessful joins due to differently named countries in both datasets
      2. To solve this, I needed to identify them and rename the countries according to the World Happiness dataset
      3. We could then perform the merge successfully!
   3. Check which ranking these Top 10 countries had in the Happiness Report and if there is a correlation there. I needed to reset the index to start at 1 in order to create a ranking and create a new column to store the ranking
   4. Once we had both rankings, I wanted to visualize the countries and both their rankings side by side in tabular format
   5. However, this analysis did not show the true picture as the number of cases and deaths are obviously relative to the country’s population, which then led to my next analysis involving a population dataset
3. Imported the population dataset to enrich my analysis capabilities involved the following steps:
   1. Cleansed the data: looking for empty data – none detected
   2. Merged the population dataset with the Covid 19 dataset using a left join on Covid19 dataset to ensure all of this data is considered and the population is added per country
   3. Once we had our merged dataset, we needed to create two new columns to show the percentage of cases and deaths per country’s population
   4. We then created a subset of the Top10 for these categories and created a new Covid Ranking for this new subset
   5. We could then merge once again with the Happiness dataset to see the new results: Country, Covid Ranking and Country Happiness Ranking
   6. In order to view this visually, I created a function to highlight the rows in either green or red depending if the Happiness Ranking was below or more than 20
   7. Portugal, Slovenia and Slovakia were all potentially impacted by Covid as they had a high Covid ranking but low Happiness ranking

**Datasets for Analysis Part 3 – World Happiness Report 2022 & Cities with the Best Work Life Balance 2022**

As my final analysis, I was curious about how the new way of working has impacted the World Happiness results. Therefore I decided to import this dataset to analyse along with the World Happiness Report. The following steps were performed:

1. Cleansed the data and created a subset of the data for my analysis (Country, City, Ranking)
2. Renamed Column Country to Country name and 2022 column to Work life Balance 22
3. Merged the work life balance dataset with the Happiness dataset and checked the top 10 of each to see if there is any correlation between the datasets
4. Switched the analysis and displayed the top Work Life Balance scores and their corresponding countries
5. Plotted using Seaborn barplots

Finally I wanted to check the impact of working-from-home on these cities by carrying out the following steps

1. added the column ‘Remote jobs’
2. Then I wanted to show the impact per country, which required a mean percentage per country calculation, as the data is per city
3. I merged with the dataset above to see how the overall work life balance rankings match with the working from home rankings
4. Some cities were ranked high in the ‘Remote jobs’ category did not appear in the top 20 in the overall Work Life Balance report

**Machine Learning potential on these Datasets**

Further Analysis could be carried out on the World Happiness Report in the area of Machine Learning. This could be carried out using the classification method which is making predictions on a categorical dependent variables and it is supervised learning as it is a labeled dataset. I would then build a model from our data by training the model. We would create a new dataset based on our data using the train-test split(80% for training, 20% for testing). Once our model is validated, we could use this to predict future results of the World Happiness report.

# Results

**Datasets for Analysis Part I – World Happiness Report 2022 vs 2021**

Chart 1:Top 10 countries per category included in the World Happiness Report 2022

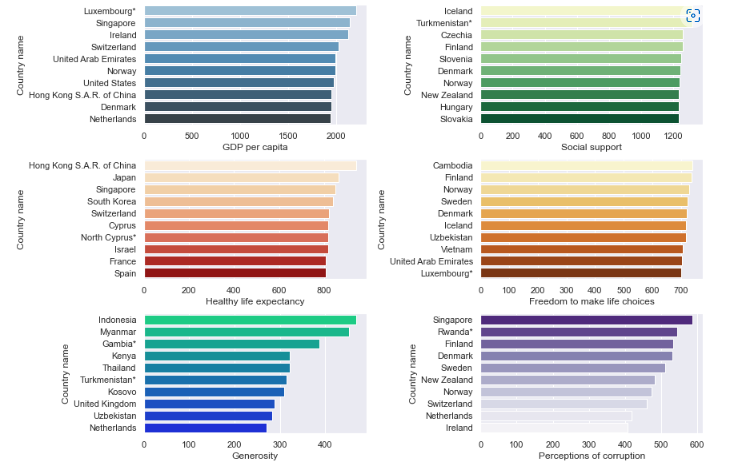


Chart 2: Happiest 10 countries vs Least Happiest 10 countries based on their ranking from the report

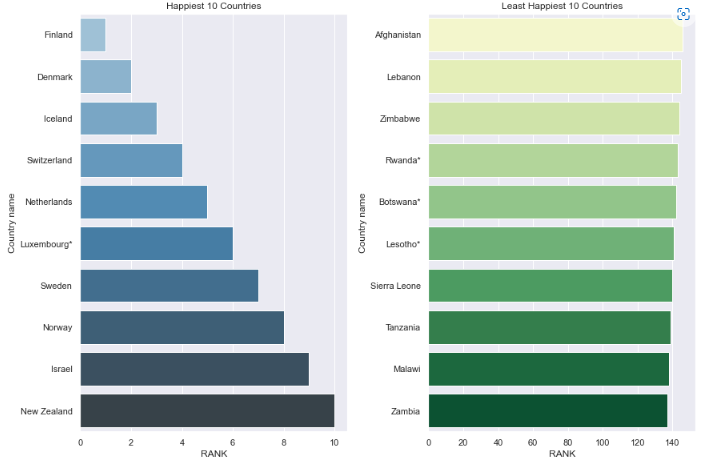


Chart 3: Comparison between Finland (number 1 rank in the world and Europe) and Ireland (9 in Europe and 13 in the world) in the two categories with the largest variance

# 

Chart 4: Comparison between Top 10 Happiest countries for 2021 vs Top 10 Happiest countries for 2022 based on their rank

# 

**Datasets for Analysis Part 2 – World Happiness Report 2022 & Covid19 API & Population Dataset**

Chart 1: Comparison between Top 10 countries for Covid Cases and Deaths, descending

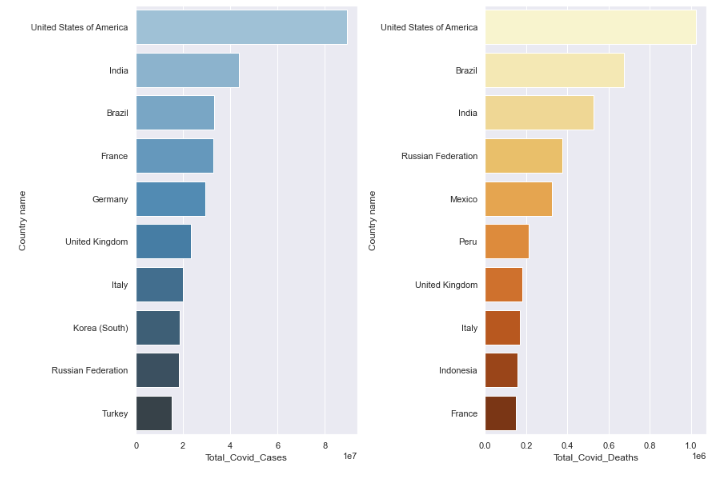


Chart 2: The comparison between Top 10 Happiest Countries alongside the Top Covid Ranking (Deaths & Cases) (red highlighted for Happiness Ranking > 20 and green <= 20, white not included in Happiness Report)

# 

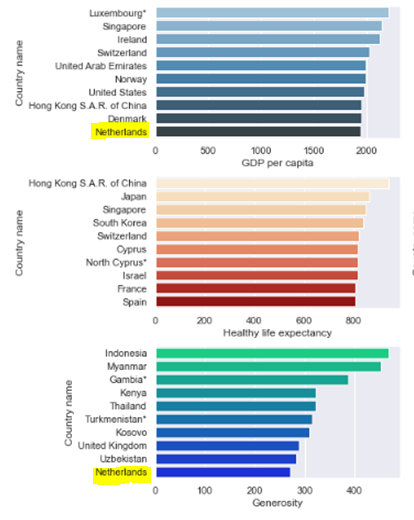
**Datasets for Analysis Part 3 – World Happiness Report 2022 & Cities with the Best Work Life Balance 2022**

Chart 1: Comparison between Top 20 Happiest Countries alongside the Most Balanced 20 cities and their respective Countries, based on rank ascending

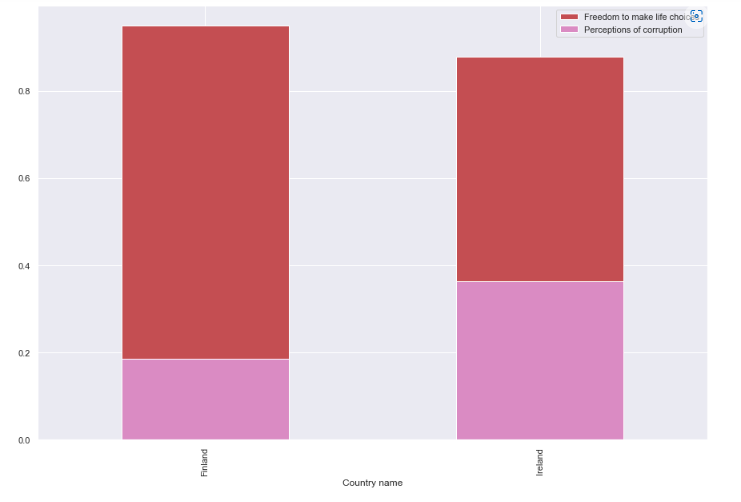
# 

# Insights

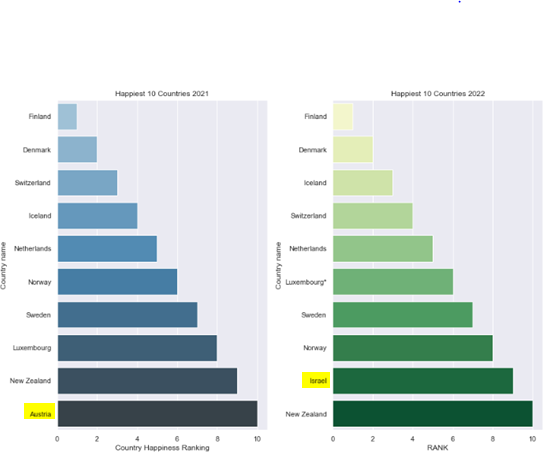
1. Upon examining the correlation between Generosity and GDP per capita, only one country in the top 10 for Generosity (Netherlands) appears in the top 10 for GDP per capita. Therefore we can derive that Generosity is not based on the wealth of a country. See highlighted below:



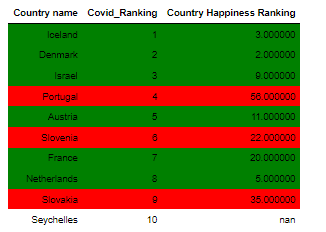
1. According to the World Happiness Report 2022, Ireland ranks number 13 in the world and 9 in Europe. Upon comparison to number 1 in the world and Europe, Finland, the following are the main categories where it falls down: ‘Perceptions of corruption and Freedom to make life choices’. See below:



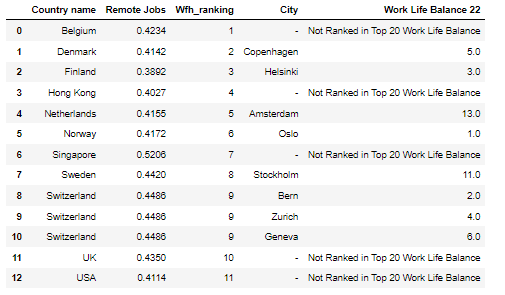
1. Upon comparing the World Happiness Report for 2022 with 2021 we can see that Austria lost their spot in the top 10 and Israel joined the ranks at number 9 for 2022:



1. Considering the Covid19 analysis, despite the high score of 3 in the World Happiness ranking, Iceland scored number 1 in Covid ranking of deaths and cases, therefore we can decipher that Covid did not impact the country’s happiness considerably. Countries such as Portugal, Slovenia and Slovakia all ranked in the top 10 for Covid cases and deaths and had quite poor happiness rankings. This could indicate an impact of covid on happiness in these countries



1. From the Working from Home vs Work Life Balance analysis results we see that wfh is not the main factor influencing Work Life Balance as 5 cities in the top wfh ranking are not in the top 20 work life balance



# References

Wikipedia World Happiness Report Available at:

<https://en.wikipedia.org/wiki/World_Happiness_Report#2022_World_Happiness_Report>

Kaggle World Happiness Report up to 2022 Available at:

<https://www.kaggle.com/datasets/mathurinache/world-happiness-report?select=2022.csv>

Kaggle World Happiness Report up to 2021 Available at:

<https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2021>

UCD Professional Acadamy Certificate in Data Analytics Available at:

[CIDA 2022-05-17 6:30pm: Jupyter Notebooks (ucdpa.ie)](https://learn.ucdpa.ie/mod/folder/view.php?id=44385) and the API:

<https://api.covid19api.com/summary>

Kaggle Countries in the World Population 2022 Available at:

<https://www.kaggle.com/datasets/anandhuh/countries-in-the-world-by-population-2022>

Kaggle Cities with the Best Work-Life Balance 2022 Available at:

<https://www.kaggle.com/datasets/prasertk/cities-with-the-best-worklife-balance-2022>