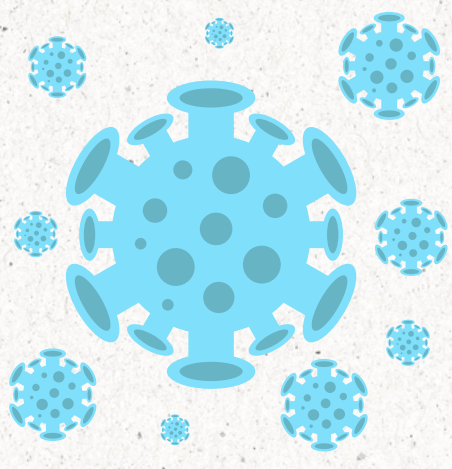
Dataset Nutrition Label

WORLD HAPPINESS

Thao Nguyen & Hyein Hwang

INTRODUCTION

Our team analyze how the happiness index of countries changed from 2015 to 2023. Along with that, we also look at suicide rates. Our team seeks to show if COVID-19 affects the world positively or negatively and if happiness scores affect suicide rates. To address those issues, my team intends to use a line chart, bar chart, and world map.



OVERVIEW

this dataset contains information about the 2018 and 2022 world happiness levels and suicide rates. On top of that, compare the happiness levels before and after the Covid-19.

COMPOSITION

Dataset represents countries, and those have several categories of calculating happiness levels. the data have three CSV files and contains columns such as "Country", "GDP per capita", and "Score". the dataset is self-contained, and it provides all needs.

COLLEECTION

the data is collected from Kaggle, and it is directly observable such as countries and seven criterias. Each data is collected in each representing year.

2020 happiness level

2018 happpiness level

Suicide rate

DEVELOPMENT PROCESS

We clarified the change of world happiness level before and after Covid-19 pandemic. In order to examine the change, we chose two years that are 2018: represent the year before Covid-19 and 2020: represent the year after Covid-19. We selected one of the criteria among factors of happiness. On top of that, to strengthen the proposal, we decided to add suicide rate so that we can compare the overall trends better.

PREPROCESSING/CLEANING/LABELING

- **Was any preprocessing/cleaning/labeling of the data done?**
- Yes, all the preprocessing/cleaning/labeling steps are done. We have collected ten different csv.file data frames and cleaned them by joining them into a single CSV file 1 file. Besides, we also create additional columns about the median rate of suicide rates.

- **Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?**
- We found the raw data from kaggle so here are the links to those and I also downloaded them into the csv file on the computer.
- Links to the raw data: https://www.kaggle.com/datasets/unsdsn/world-happiness/data?select=2019.csv
- https://www.kaggle.com/datasets/londeen/world-happiness-report-2020
- https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2021?select=world-happiness-report-2021.csv
- https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2022
- https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2023
- https://www.kaggle.com/datasets/marshuu/suicide-rate-and-life-expectancy?select=Suicide+Rate.csv

- **Is the software that was used to preprocess/clean/label the data available?**
- The software we are using to preprocess/clean/label the data is R studio which is openable and readable.

USES

- **Has the dataset been used for any tasks already?**
- We have used the datasets to merge them together in 1 column and also create a new column to calculate the median of suicide rates.

- **Is there a repository that links to any or all papers or systems that use the dataset?**
- Yes, we have created a repository on GitHub. Here is the link: https://github.com/INFO-201-Fall-2023-Final/final-projects-ellenandhyein

- **What (other) tasks could the dataset be used for?**
- The datasets will also be used for visualizing the data to become line, bar and world map. In order to create those charts, we need to based on the datasets to have the information and numbers.
- Firstly, for the column chart, we will compare the happiness values of the top 5 happiest countries between 3 periods of time: before, during, and after COVID.
- For the bar chart, it is the combination between bar and line chart. For the bar, it is for suicide rate while the line is for the happiness score. By doing this, it highlights the relevance between those two components.
- Next, regarding the bar charts, we will show the happiness score of the top 50 countries by using a movable selection for users to interact with our website.
- We will create the second line chart to show how 7 factors contribute to the happiness score. More specifically, my team will point out which factor contributes the most to the score.
- Lastly, our team will do a world map. For this kind of visualization, people will be able to see the suicide rate of all countries in the world in 2020.

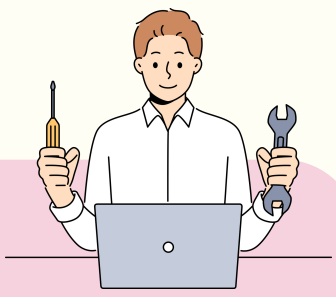
- **Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?**
- The only thing that users need to notice is that make sure to pay attention to the guidelines that will be shown when you enter the websites. Reading carefully the description on how to use the graphs is the best way to have the best experience and understanding while using graphs.

- **Are there tasks for which the dataset should not be used?**
- All the datasets that we have found will be used effectively in this project.

DISTRIBUTION

- **Will the dataset be distributed to third parties outside of the entity on behalf of which the dataset was created?**
  - No, it will not be distributed to third parties.
- **How will the dataset be distributed? Does the dataset have a digital object identifier (DOI)?**
  - The dataset will be distributed on GitHub. No, it does not have a DOI.
- **When will the dataset be distributed?**
  - The dataset will be distributed by the end of this quarter.
- **Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?**
  - No, the dataset will not be distributed.
- **Have any third parties imposed IP-based or other restrictions on the data associated with the instances?**
  - No, there are no third parties imposed IP based or other restrictions.
- **Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?**
  - No, there are no export controls or other regulatory restrictions apply to the datasets or to individual instances.

Maintenance

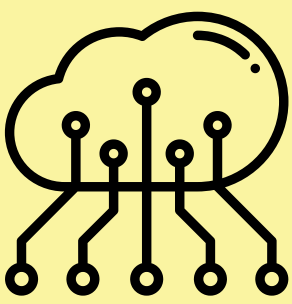


- **Who will be supporting/hosting/maintaining the dataset?**
- All members of the team will be in charge of that.
- **How can the owner/curator/manager of the dataset be contacted ?**
- Here are ways to contact our team through gmail: thao19@uw.edu and hyeinh@uw.edu .
- **Is there an erratum?** No, there is no erratum.
- **Will the dataset be updated?**
- This dataset is still in process so we are still continuing editing it in order to have a completed version. However, in the future, when it needs to be fixed, our team will always be willing to have the dataset be updated. The updates will be announced through GitHub.
- **If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances?**
- No, there are no applicable limits on the retention of the data associated with the instances.
- **Will older versions of the dataset continue to be supported/hosted/maintained?**
- Whenever the dataset has any errors and information needs to be updated, our team will fix it as soon as possible based on the older versions. The new updates will be announced through GitHub and we will also highlight which parts are updated in order for the consumers easily following.
- **If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? Will these contributions be validated/verified? Is there a process for communicating/distributing these contributions to dataset consumers?**
- Every comment or input is appreciated by our team. To allow build-ons to our datasets, we will provide a form on GitHub. Replies will be processed in 72 hours and updates posted on GitHub. We value these inputs that will change perspectives. It may help our staff improve our datasets. We will carefully review all feedback. However, not every submission is approved and used for updates because we must employ numerous criteria to determine if they are suitable for our databases. However, we must reiterate that every opinion is valued and we welcome all contributions.

Impacts and Challenges

This dataset may exhibit confirmation bias as most of the countries were shut down and people had many restrictions during Covid-19 that the result can show like the happiness level will drop after Covid-19.

Collection Process



The data is directly observable such as countries and seven criteria. Since all the data are from Kaggle, and there are no explanations about how they collected them. There are no ethical review processes.

Acknowledgements

We thank Thao Nguyen and Hyein Hwang as being the greatest teammates in the team who collaborated and worked together perfectly to complete this dataset. Besides, we also feel thankful for Henry Ramstad, Julia Deeb-Swihart, and other instructors who have given feedback and instructions to our work.