

# Impact of Covid-19 on World Happiness

Ho Hang Cheng

**Abstract**—This is the forth year since Covid-19. During the pandemic people had came through a huge changes in society. This study aims to find out the impact of Covid-19 on people happiness. We would use time series analysis, clustering and correlation heatmap to help solving our research questions. The happiness index have no significant changes over the year, but we found that the patterns in happiness index has changed among countries. Some factors become more important to the happiness score.

---

## 1 PROBLEM STATEMENT

Since Covid-19 was declared a pandemic by the World Health Organization on March 2020 [1], society forms in countries had been reformed by different restrictions published by governments. Besides the fear of infections, people were facing arguments about vaccine safety, travel restrictions, lockdowns, social distancing etc. The pandemic has forced us all to drastically change our daily habits, our thoughts on hygiene and how we interact with each other in a short amount of time. Isolation from friends and family had been increased due to social distancing rules.

To discover the impact of covid-19 on human activities, The World Happiness Report would be the indicator to find out the changes in the world before and after the pandemic. The World Happiness Report, a publication of the Sustainable Development Solutions Network, is powered by Gallup World Poll data which shows the ranking and analysis among 150 selected countries base on their happiness index [2].

This study aims to answer the following research questions:

1. What is the difference between world happiness before and after covid-19?
2. Has covid-19 changed which factors determined happiness score the most?

The selected dataset is capable to answer the questions as it consists of world happiness data from 2015 to 2022, which includes the newest happiness score of selected 146 countries. It allows us to make a comparison on happiness score before and after Covid-19.

## 2 STATE OF THE ART

The first World Happiness Report was released in 2012 and has since been published each year. It contains an analysis of the state of global happiness and provides information, rankings, and policy recommendations on how countries can improve their well-being. The report focuses on factors such as health, social support networks, freedom, government corruption and GDP per capita. It is meant to provide governments and policymakers with empirical evidence on the efficacy of various policies around the world in improving their citizens' sense of well-being. Last year is the 10<sup>th</sup> Anniversary of the World Happiness Report, it is still a new domain in our world.

From the two research [3][4] we have analysed, both applied visual analytics on solving their research questions about world happiness. For [3], not only World Happiness Report but also datasets that are related to happiness and life satisfaction were used in analysing the phenomenon of happiness, such as the World Value Survey and Public Opinion surveys. They studied the correlation among the factors, for instance, life expectancy, income, and sense of freedom. They built several interactive charts such as happiness scores over time, the correlation between happiness score and life expectancy among countries by scatter plots and maps showing countries' happiness scores. Their purpose is to solve the questions about how people are satisfied with their lives in different countries, were people happier in the past and how our living conditions affect all of us [3]. Regarding the visualisations they have made, they found that in general, the wealthier the country, the higher in happiness score, and European countries top the ranking (Finland, Denmark, Iceland, Switzerland and Netherlands).

For [4], Their work is to use visualisations to make users can easily understand the phenomenon and trends of World Happiness. They used The World Happiness Report as well as their data. To let any users able to understand and interact with the indexes and answer different user questions, they have five tasks to do in their work: present the features, discover happiness distribution, identify happiness score by countries, locate any country that the user wants to find, identify the extreme and discover happiness distribution in regions [4]. They have made a choropleth map that can express the happiness score and other features among countries and can manipulate by users.

Both pieces of research showed how the world happiness can be expressed in different forms and give insights into our study. What we have learned from them is how we can interpret world happiness such as comparing features by their correlations with happiness scores and the idea of their visualisations.

## 3 PROPERTIES OF THE DATA

In the World Happiness Report, there are 6 factors to determine the happiness score of a country. They are the categories help account for the differences in life satisfaction around the world: GDP per capita, Healthy life expectancy, Freedom to make life choices, Social support, Generosity and Absence of corruption. The report uses data from the Gallup

World Poll surveys. They are based on responses to key life assessment questions posed in the polls by the method of Cantril ladder [5]. Respondents are asked to think of a ladder, with the best possible life for them being a 10 and the worst possible life being a 0 [6]. The sample size in each country is 1,000 people per year.

The dataset is retrieved from Kaggle.com. It contains 8 years data of the World Happiness Report from 2015 to 2022. Besides the 6 factors and happiness scores, happiness rank, country, region, year and index of each sample are included in the dataset. There are 12 columns and 1,229 rows in total.

We used histogram and boxplot to show the quality of the data. Figure 1 was showing the distributions of the 6 factors by histograms. We can see that the GDP per capita is showing a normal distribution. For Social Support, Life Expectancy and Freedom, their distributions are slightly skewed into right, but they are still performing a bell-shape. For Government Corruption and Generosity, their distributions were skewed into left.

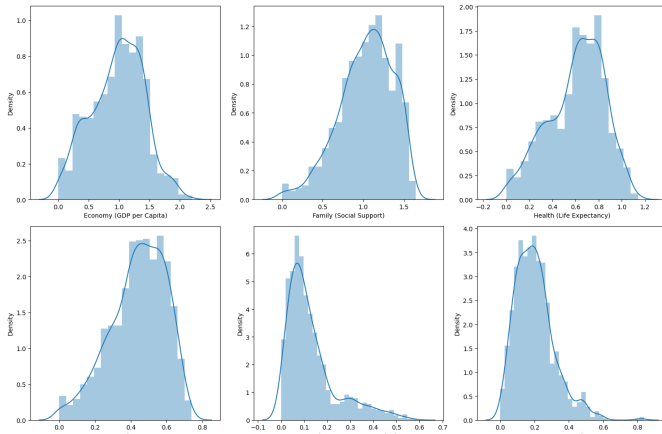


Fig. 1. Distributions of factors in histogram

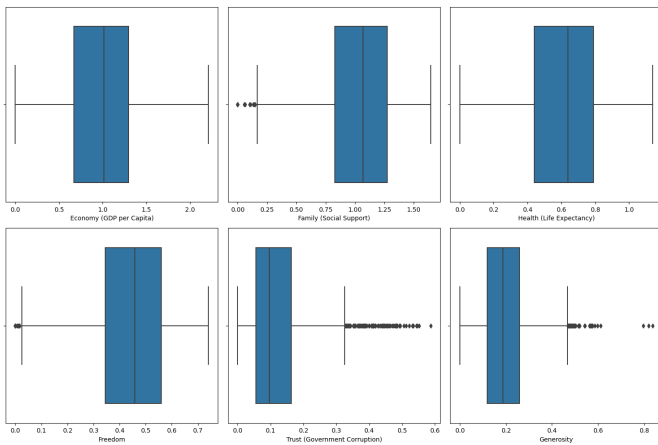


Fig. 2. Boxplots of factors

In figure 2, the boxplots were showing the quartiles and the outliers of the 6 factors in the dataset. In this case, we didn't remove the outliers, because in this dataset each sample is

representing its own country, any rows are meaningful to this study.

We removed two columns in the dataset which are the row index and the happiness rank of countries, as they wouldn't be useful in our study. There is no missing value in the original dataset, it was cleaned by the dataset provider in Kaggle.com. The dataset was spitted into two parts: the data from 2015 to 2019 and the data from 2020 to 2022. It would make us easier to make visualisation in the analysis part. We have also checked the distribution on both set of data, and they were showing similar patterns as the original dataset.

## 4 ANALYSIS

### 4.1 Approach

To answer our research questions, we want to make a comparison of the data before Covid-19 and data after Covid-19 by different methods of visualizations. The first step in our analysis part is data pre-processing. We divided the original dataset into two parts and based on these two datasets made data wrangling and visualization. This step consists of human reasoning as this decision is based on our knowledge about when Covid-19 declared a pandemic, and the process was done by computer. We would perform three tasks in the analysis part to solve the two research questions.

#### Task 1

Task 1 and Task 2 would mainly use for solving the first research question: What is the difference between world happiness before and after covid-19. We cannot simply solve this question by one or two visualizations as we want to find out all the differences in happiness score between before and after Covid-19, including the changes in the happiness factors. The first task we would carry out a time series analysis to see the changes in happiness score and the factors over time. The human judgement in this task is to consider the hue of the graphs. We can use country or region to see which one is better in performing visualizations.

#### Task 2

Task 2 we would use the K-means clustering algorithm [7] to assign countries into different clusters based on their happiness factors, and we would perform it on both datasets to see if there are any changes in the two outcomes. To pick the optimal value of "K", the Silhouette score method [8] would be applied. We would visualize Silhouette plots and decide the optimal "K" by human judgement. In the end we would use choropleth maps to show the clusters of countries, it would be easier for the user to find out the patterns.

#### Task 3

For our second question: Has covid-19 changed which factors determined happiness score the most, we would compare the correlation between the happiness score and the factors, on the two separated data. We would use scatter plot and correlation heatmap to show the changes. The correlation

can tell us which factors have the most influences on the happiness score. We would also see if the factor with the highest correlation in the 2015-2019 dataset will also have the same place in the 2020-2022 dataset.

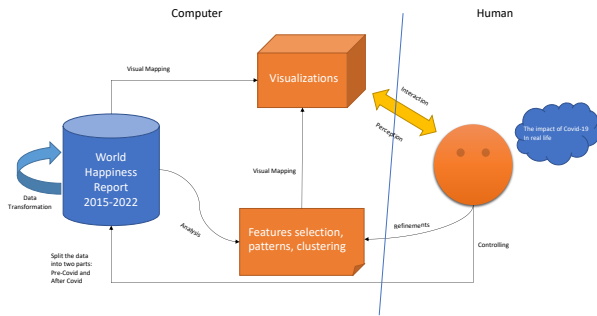


Fig. 3. Workflow diagram

## 4.2 Process

### Task 1

We used Python to handle all the computer processes in our study, including data pre-processing, data transformation and visualization. In the first task, we performed data transformation at first: group by the data by years and taking the average of each factor to make a time series analysis in figure 4.

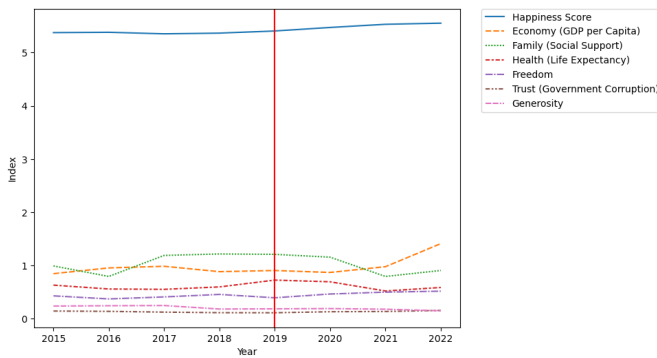


Fig. 4. Time Series Analysis in Happiness Indexes

From the graph, we can see that a red axis was added in 2019 to help differentiate the indexes before and after Covid-19. Overall, there were no significant changes in the happiness indexes over the years, all the lines were showing a flatter trend. More specifically, the average happiness score slightly increased from 2019 to 2022. For the factors, the Social Support index dropped to the 2016 level in 2021, it had switched the position with the GDP per capita. Compare with the value before 2019, the average GDP per capita rises to a new high level in 2022. Besides these two factors, Healthy Life Expectancy is also worthy of attention. Although the change was small, it started dropping in 2019. For Freedom to make life choices, it had a slight growth from 2019 to 2022. The other factors didn't have significant changes.

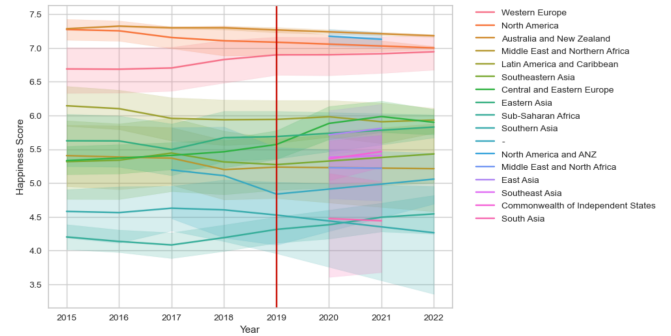


Fig. 5. Time Series Analysis in Regions

We had also taken the average happiness score by region to see any specific changes over the year in figure 5. Instead of countries, we used regions as the hue as there were 195 unique values in the "Country" column of the dataset, the graph would be messy if we used it. There are 17 unique values in the "Regions" column in total over the years. As well as figure 4, we didn't see any significant changes in happiness scores among the regions. Although we couldn't see the difference from figure 5, the visualization informed us which regions had higher happiness scores over the years. Australia and New Zealand are the region that got the highest happiness score on average from 2015 to 2022.

### Task 2

As we mentioned in our approach, we would carry out the K-means clustering to assign countries into "K" clusters. It would be performed in both datasets to see the changes of clusters among the countries. To decide the optimal value of "K", we used the Silhouette plots in figure 6 to help us make the human judgement.

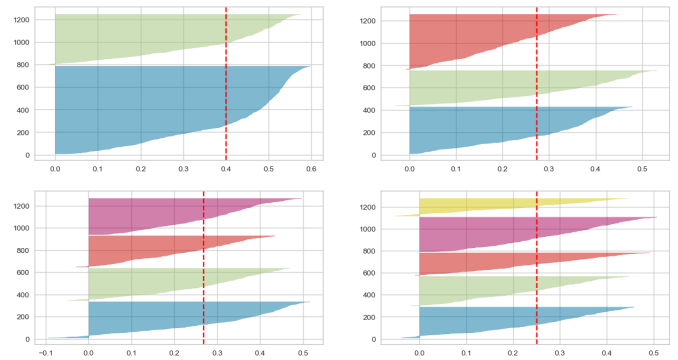


Fig. 6. Silhouette plots for K = 2 to 5

To pick the best value of "K" by using the Silhouette plots, we should check two things: are the Silhouette scorer of all the clusters higher than the average score of the dataset, and the average score is represented by a red axis in the plots. Second is there should not be wide fluctuations in the size of the clusters. The width of the clusters in the plots represents the number of data points [9]. From figure 6, all the cluster had a higher score than the average, so, we determined K by the average size of the clusters. For K = 2 to 3, one of their clusters was larger than the others. In K = 4, the size of the four clusters

were balanced. Regarding to our observation, the value of K would be 4.

Then we moved to the computer section to perform the K-means clustering. We used KMeans function from SciKit-learn [10] to do the task in Python. After assigned clusters to each country in both datasets, we visualized the results by the choropleth maps.

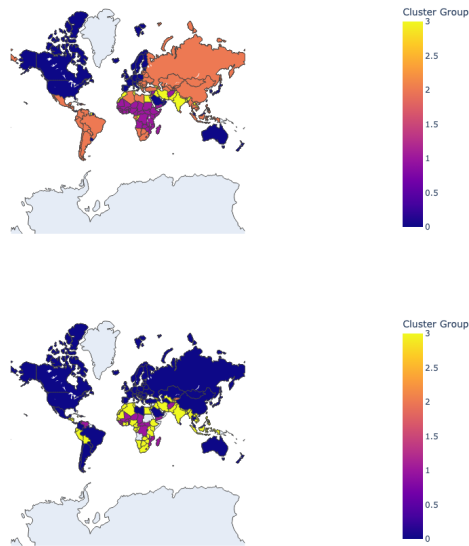


Fig. 7. Choropleth maps for 2015-2019 and 2020-2022

From figure 7, we could see that the pattern of clusters has changed. In the upper figure, the map is representing the world before Covid-19. The United States, Canada, Western Europe and Australia were in the same cluster. Russia, China, Mexico and countries in South America were in the cluster 2. The countries in the same cluster were having similarity in their happiness index. In the lower figure, most of the countries that belonged to cluster 2 in figure 7 changed into cluster 0. In overall the happiness index among countries becoming similar, cluster 0 became the majority.

Task 3

In this section, we would use scatter plots and correlation heatmaps to show the changes in correlation between happiness score and the factors. From the correlations, we can see which factors have a stronger positive relationship with happiness score, and we want to find out would the factors with the highest correlation in both datasets be the same factor.

Figure 8 was showing the correlations by scatter plots. The upper sixth were representing data from 2015 to 2019 and the lower sixth were representing 2020 to 2022. For 2015 to 2019, the factor with the highest correlation with happiness score is GDP per capita, the Pearson's correlation coefficient is 0.79, followed by Life Expectancy (0.74) and Social Support (0.65). Generosity has the lowest positive correlation with happiness score, which has only 0.14.

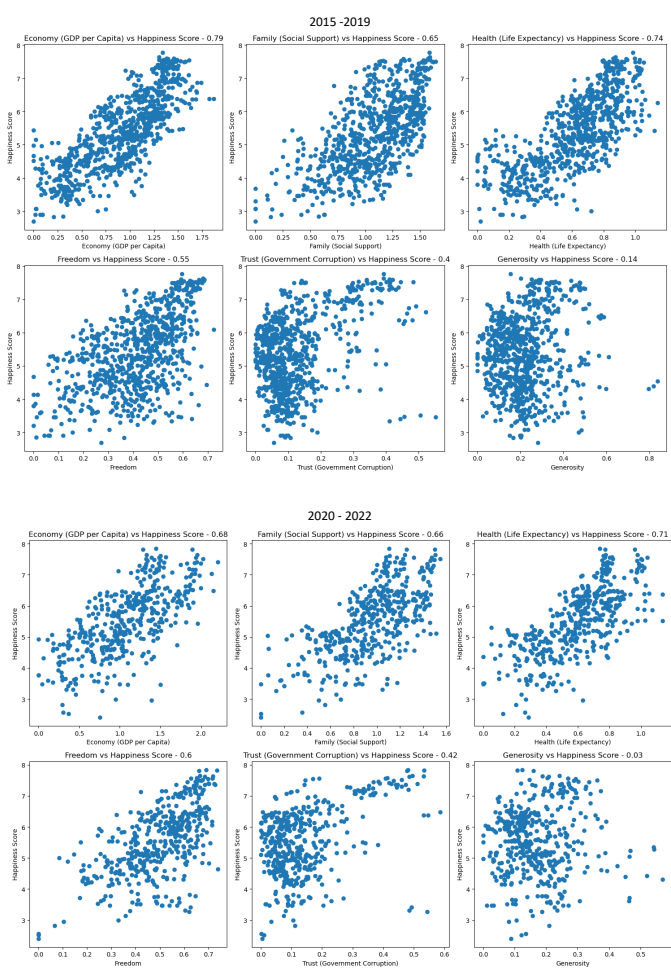


Fig. 8. Scatter plots: happiness score vs factors

In 2020 – 2022's figures, GDP per capita was no longer the factor that has the strongest positive correlation with happiness score. Life Expectancy became the factor with the highest correlation coefficient which was 0.71. Generosity was still the factor that had the lowest correlation with happiness score.

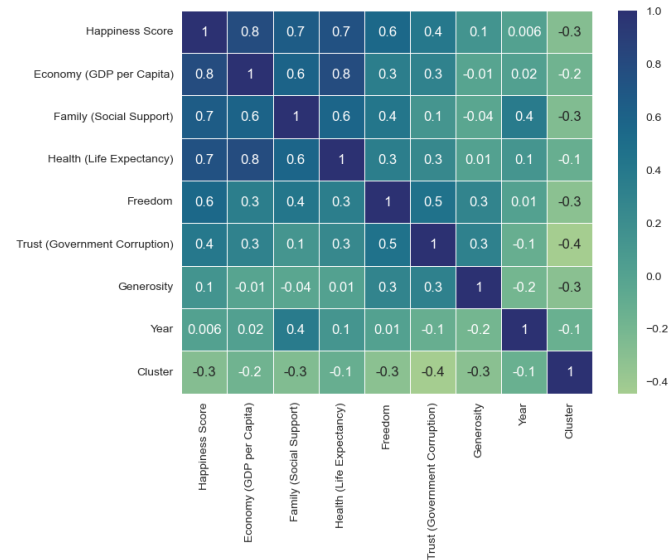


Fig. 9. Correlation Heatmap for 2015 - 2019



Fig. 10. Correlation Heatmap for 2020 – 2022

Figures 9 and 10 are the correlation heatmaps for numeric features in the dataset. The rise of Life Expectancy can be explained by awareness of health have been increased since the Covid-19. Besides GDP per capita and Life Expectancy, Social Support also had strong positive correlation with happiness score, and by comparing the two heatmaps, the correlation coefficient of Social Support was increased and had the same importance as the first two factors. During pandemic, social support provided by governments became more important than usual as people were facing isolation and less contact with each other's.

### 4.3 Results

From the time series analysis in task 1, we discovered there was no significant changes in happiness score over time, even after the pandemic. The average happiness score slightly raised which is unexpected. The wealthy regions such as North America and Western Europe remain high in happiness score over the time, Covid-19 didn't bring big impact on them. In task 2, we found that the big countries such as the United States, Russia and China were into the same cluster in 2020 – 2022 dataset, their happiness factors were similar. Only Africa remains different cluster with the other continent when comparing the two choropleth maps. The above are the answer to our first research question.

The second question is answered by task 3. We found that Healthy Life Expectancy became the most important factor among the happiness indexes. The Social Support also became more important than before the pandemic. Despite the rise of average GDP per capita in task 1, the correlation of it and happiness became weaker after the pandemic.

### 5 CRITICAL REFLECTION

In conclusion, we are capable to answer the research questions by the selected dataset, but there are still some improvements can be made in our analysis part. For the time series analysis, it is a good method to easily show the happiness

indexes over the years, we could find the trend inside the graph quickly. In figure 5, one of the lines was representing '-', it was because some countries can't specify into any regions, and 17 lines in a graph is quite hard to see changes in a particular line. To advance the method in further study, we can engineer the features, simplify countries into less regions and make cleaner graphs.

The choropleth maps gave big help in solving the first research question, as we couldn't find significant changes through the time series analysis. By the maps we found that some countries in the same cluster in 2015-2019 were switched into another one in 2020 – 2022. This tell us the pattern in happiness index are different in two set of data. But one thing is that the data representing world happiness in Covid-19 only consists of 3 years data, which is imbalance with the 2015 to 2019 data. In further study, we suggest collect more data in the coming years, to make the comparison become fairer.

Although the dataset can solve our research questions, we have also restricted by it. The World Happiness Report only consists of 6 factor to determine the happiness score of a country, despite the other meaningful data such as salary, house prices, land area of the countries etc. Regards to [3], their work consists of number of datasets, which can provide more possibility in analysing the world happiness and make more specific comparisons.

The limitation in our study is we cannot make a more specific analysis in a single country like comparing happiness between towns in a country, due to our research questions we can only make a macro-observation on World Happiness Report.

Our recommendation for analysts having to deal with similar data is collect more data from different sources, it would help make the area of the research deeper, and try more visualization tools such as Tableau.

### Table of word counts

Problem statement	229/250
State of the art	433/500
Properties of the data	393/500
Analysis: Approach	391/500
Analysis: Process	994/1500
Analysis: Results	170/200
Critical reflection	364/500

### REFERENCES

- [1] Brannen, S., Ahmed, H., & Newton, H. (2020). *Covid-19 Reshapes the Future*. Center for Strategic and International Studies (CSIS). <http://www.jstor.org/stable/resrep25198>
- [2] Helliwell, J. F., Layard, R., & Sachs, J. D. (2012, April 4). *World happiness report 2012*. The World Happiness Report. Retrieved from <https://worldhappiness.report/ed/2012/>
- [3] Esteban Ortiz-Ospina and Max Roser (2013) - "Happiness and Life Satisfaction". Published online at OurWorldInData.org.



Retrieved from: 'https://ourworldindata.org/happiness-and-life-satisfaction' [Online Resource]

- [4] Bazurto-Gomez, N. *et al.* (2019). An Information Visualization Application Case to Understand the World Happiness Report. In: Agredo-Delgado, V., Ruiz, P. (eds) Human-Computer Interaction. HCI-COLLAB 2018. Communications in Computer and Information Science, vol 847. Springer, Cham. [https://doi.org/10.1007/978-3-030-05270-6\\_4](https://doi.org/10.1007/978-3-030-05270-6_4)
- [5] Levin, K.A., Currie, C. (2014) Reliability and Validity of an Adapted Version of the Cantril Ladder for Use with Adolescent Samples. *Soc Indic Res* **119**, 1047–1063. <https://doi.org/10.1007/s11205-013-0507-4>
- [6] *FAQ*. (n.d.). Retrieved from <https://worldhappiness.report/faq/>
- [7] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108. <https://doi.org/10.2307/2346830>
- [8] K. R. Shahapure and C. Nicholas (2020), "Cluster Quality Analysis Using Silhouette Score," 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 2020, pp. 747-748, doi: 10.1109/DSAA49011.2020.00096.
- [9] Tomar, A. (2022). *Stop using elbow method in k-means clustering, instead, use this!* Medium. Retrieved from <https://towardsdatascience.com/elbow-method-is-not-sufficient-to-find-best-k-in-k-means-clustering-fc820da0631d>
- [10] *Sklearn.cluster.kmeans*. scikit. (n.d.). Retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
- [11] Annabel, Kay. (n.d.). *The Science of Happiness – A Visual Exploration*. Retrieved from [https://moodle.city.ac.uk/pluginfile.php/2808074/mod\\_resource/content/1/117002361-166227\\_Annabel\\_Kay\\_Individual\\_Project\\_-\\_Annabel\\_Kay\\_1418814\\_906946061.pdf](https://moodle.city.ac.uk/pluginfile.php/2808074/mod_resource/content/1/117002361-166227_Annabel_Kay_Individual_Project_-_Annabel_Kay_1418814_906946061.pdf)

## DATA

*World Happiness Report 2015 to 2022.*

Retrieved from

<https://www.kaggle.com/datasets/mayzannilarthain44/world-happiness-report-2015-to-2022>