Exploratory Data Analysis on Healthy Lifestyle & World Happiness Report in Asia & Europe

Karnapali Mukhopadhyay [1], Chandrani Adhikary[2], Dola Saha [3], Chiranjib Dutta [4], Rooprekha Baksi [5]

[1] Department of Computer Applications
Guru Nanak Institute of Technology
157/F Nilgunj Road, Panihati, Kolkata – 700114, West Bengal

[2]Deptartment of Computer Applications
Guru Nanak Institute of Technology
157/F Nilgunj Road, Panihati, Kolkata – 700114, West Bengal

[3]Assistant Professor, Dept. of Computer Applications
Guru Nanak Institute of Technology
157/F Nilgunj Road, Panihati, Kolkata – 700114, West Bengal

[4]Assistant Professor, Dept. of Computer Applications
Guru Nanak Institute of Technology
157/F Nilgunj Road, Panihati, Kolkata – 700114, West Bengal

[5]Assistant Professor, Amity Institute of Psychology and Allied Sciences
Amity University Kolkata
Corresponding author's email: karnapalimukhopadhyay1@gmail.com

Abstract

State of mind and lifestyle have been affected due to COVID -19 pandemic. There has been a certain impact on daily lifestyles. There is one the metrics like happiness index, to be one portion to exhibits that each country has an indicator which affects each other. This indicator is required to understand the impact of the happiness index and lifestyles. In the dataset, it is necessary to explore, analyze, and visualize to give knowledge about the citizens of Asia & Europe. Exploratory data analysis (EDA) has been utilized for further investigations. These works explore happiness reports and lifestyles in the Asia & Europe region using a combined dataset of World Happiness Report 2020 & Healthy Lifestyle report 2021. The result describes and discusses the datasets using a table with correlation score, correlation analysis, and heat map-plotting visualization. In the future, this study will further investigate different regions with artificial intelligence models.

Keywords

 Exploratory Data Analysis; Healthy Lifestyle report; Happiness Reports; Data Science; Visualization; Correlation analysis;

1.0 Introduction

WHO recommends a healthy lifestyle that includes eating plenty of fruits and vegetables, limiting fat, sugar, and salt intake, and exercising. People can check their body mass index (BMI) based on their height and weight to see if they are overweight. WHO publishes a variety of materials to promote and support healthy living.

Exploratory Data Analysis (EDA) is part of the data science process. It's very important before doing feature engineering and modelling the data since at this point, we have to grasp the typical data first. Data science process have many steps, there are data preparation, data cleansing, exploratory data analysis, feature engineering, data modeling, data evaluation, and deployment. In this scenario, perhaps EDA was used to knowing the contents of data from distribution data, frequency, correlation, and the others. In the experiment, curiosity is more significant in this instance, to grasping the main idea of the data need to verify, because it can answer the deep of the problem.

In this works, when the data is acquired, then for the next step is it is examined. Analysis data is the process of arranging the sequences of data, organise them into a pattern, category, and basic description units. Based on the fundamental concept, data analysis work contains three processes namely preparation, tabulation, and application of the data appropriately with a research method that is processing data by applying formulas or rules existing according to the research approach. Hence, we can infer that data analysis is a process arrange data, perform synthesis, processing data using formulas or the rules according to the approach and draw conclusions.

Before the data is turned into data modelling, the data should be evaluated using exploratory data analysis. Certainly, in this research the data is acquired from Kaggle using data World Happiness Report & Healthy Lifestyle Report. The includes of the statistics there are about the country, regional indicator, social support, GDP, and etc. all. In the other hand, for this research will be emphasis in the Comparative data analysis at Asia & Europe.
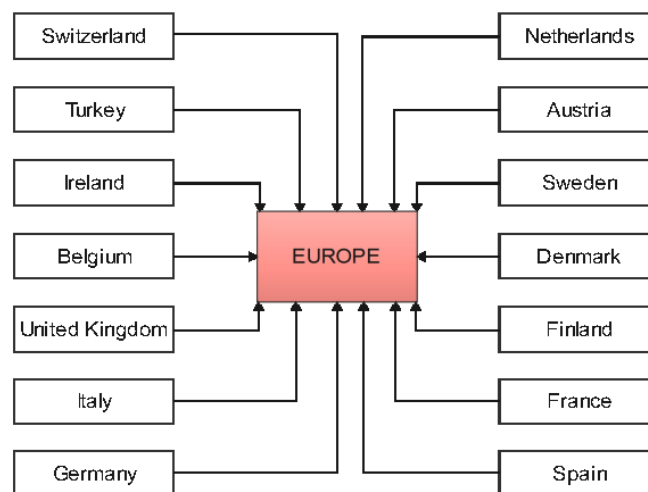

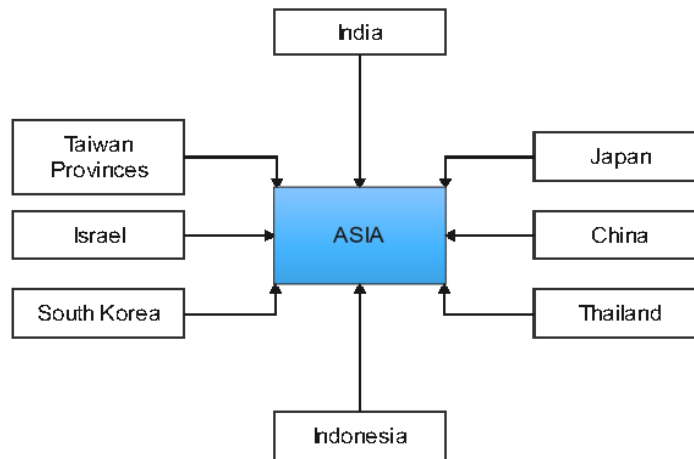Fig1. Dataset obtained from different countries of the Europe

Fig2. Dataset obtained from different countries of the Asia

Happiness Index

The happiness index is a tool for the use of researchers, community organisers, and policy makers attempting to understand and promote individual happiness, community well-being, social justice, economic equality, and environmental sustainability. Community organisers, students, researchers, educators, government and the others that is the indicator of happiness to promote social change by producing the survey instrument. The index is unique in that it is the only widely comprehensive index available for free online that survey takers to access their own scores in comparison the entire data set, while also allowing users to customise the tool for a target population, add their own questions to the survey instrument, and readily access data for their own sample as well as comparison data to the entire data set.

The Happiness Index measures life satisfaction, the feeling of happiness, and other happiness. domains: psychological well-being, health, time balance, community, social support, education, arts, culture, environment, governance, material well-being, and work. The Happiness Index is a survey that measures your happiness in 10 different areas of our life. The data collected from the survey is used to evaluate the happiness of people.

Exploratory Data Analysis

EDA is data analysis approach. It's having three popular data analysis approach as follow classical, exploratory (EDA), and Bayesian. These three processes are similar in that they all start with a general science/engineering problem and all yield conclusions. The difference is the sequences and focus of the intermediate steps. For the first classical analysis is begin from problem, collecting data, data modelling, data analysis, and conclusion. The second EDA has sequences, begin from problem, collecting data, data analysis, data modelling, and conclusions. The last approach is Bayesian was started from problems, collecting data, data modelling, prior distribution, data analysis, and conclusions. On the contrary, the EDA technique allows the data to suggest admissible models that best suit the data. The primary goal of EDA is to maximise the analysist insight into a data set and into underlying structure of a data set, while providing all of the specific items that an analysist would want to extract from a data set, such as a good fitting, a list of outliers, a sense of robustness of conclusions, estimate for parameters, uncertainties for those estimates, a ranked list of important factors, conclusions as to whether individual factors are statistically significant, and optimal settings.

Methodology

In this study employing the exploratory data analysis (EDA) method, where EDA is an approach to extract the information enfolded in the data and summarise the key characteristic of the data

We use dataset World Happiness 2020 (csv) & Healthy Lifestyle Report 2021(csv) from Kaggle, and the content there include country, regional indicator, ladder score, ladder score standard error, upper and lower whiskers, logged GDP per capita, social support, freedom to make life choices, and etc. all. Because the researchers were lived in the Asia & Europe, consequently this works will be focus on the country of Asia & Europe.

In this section, we will describe the stages by steps to explore the data, from get the data to visualisation.

Part I, we set a as csv file and import into jupyter notebook.

Furthermore, we established various library to analysis and visualisation the data, as follow utilising numpy, pandas, matplotlib, seaborn, plotly-express, plotly-graph. After setting up the data, we change the name in many columns, why we change those, because the data need to generalise perception, so the others reader can understand of the purpose each column in the data. In the other hand, the data is sorted list depending on rank world happiness report & healthy lifestyle report.

Part II, the columns was no need to analysis is eliminated by us, so we leave a few columns to be analyzed.

Here, we generated the EDA report based on Healthy lifestyle & World Happiness metrics:

1.   Sunshine hours

Sunshine length, also known as sunshine hours, is a climatological indicator that represents the duration of sunshine in a certain time (usually, a day or a year) at a specific place on Earth and is commonly provided as an averaged value across many years. It is a wide indication of a location's cloudiness and hence varies from insolation, which measures the total energy provided by sunshine over a specific time.

2. Cost of a bottle of water

The price per unit of any case of alcoholic liquor, other than beer, should be established by dividing the case price by the number of units or bottles in such case price and adding an amount not less than the following to the quotient: A unit or bottle containing less than half a pint or 200 millilitres costs two cents; a unit or bottle containing more than half a pint or 200 millilitres but less than one pint or 500 millilitres costs four cents; and a unit or bottle containing more than one pint or 500 millilitres costs eight cents.

3. Obesity levels

Obesity is one side of the double burden of malnutrition, and currently more people are fat than underweight in every region except sub-Saharan Africa and Asia except Europe. Overweight and obesity, once thought to be an issue only in high-income nations, are now considerably on the increase in low- and middle-income countries, particularly in metropolitan areas. The great majority of overweight or obese youngsters are found in poor countries. where the pace of increase has been more than 30 percent higher than that of industrialised countries

4. Life expectancy(years)

Life expectancy has increased dramatically, with many industrialized countries expecting an average life expectancy of 85 years or more by 2030. A longer lifespan is associated with socioeconomic characteristics such as sociodemographic status, GDP per capita, healthcare expenditure, universal health coverage, and laws and regulations. Meanwhile, in industrialized countries, the combined influence of modifiable healthy lifestyle characteristics such as moderate physical exercise, a healthy body mass index (BMI), non-smoking status, moderate alcohol intake, and enough sleep length has been linked to an extended lifespan. This study shows that adopting a manageable healthy lifestyle can help people live longer lives.

5. Pollution (Index score)

The Comprehensive Environmental Pollution Index (CEPI) is a rational number used to characterize the quality of the environment at a specific area using the source, route, and receptor approach. A rise in the CEPI score indicates that the receiving environment is suffering.

6. Annual avg. hours worked

The total number of hours actually worked per year divided by the average number of individuals in employment per year is the average annual hours worked. Actual hours worked include comprehensive, component, and half workers' usual work hours, compensated and underpaid overtime, work hours in new jobs created, and duration not worked owing to public holiday periods, full pay leave, own disease, damage, and temporary disability, parental leave, parental leave, school education or mentoring, idle time work for technical or financial purposes, blow or labour dispute, poor weather, remuneration leave, and other reasons. The data include both employees and self-employed folks. This amount is reported in hours per worker per year. The following health warning is issued with the data: The data are aimed for comparing trends over time; however, owing to changes in their sources and methods of production, they are unsuitable for comparing the quantity of average annual hours of work for a single year.

7. Happiness levels

In the context of mental or emotional states, happiness refers to good or pleasant emotions ranging from satisfaction to profound delight. Life satisfaction, well-being, subjective well-being, flourishing, and eudaimonia are some of the other types.

8. Outdoor activities

Outdoor recreation or outdoor activity refers to recreation done outside, most typically in natural environments. The activities that involve outdoor recreation differ based on the physical location they are being carried out in. These activities can include fishing, hunting, camping, and horseback riding – and can be accomplished individually or collectively. Outdoor recreation is a broad notion that spans a varying variety of activities and settings.

9. Number of take-out places

prepared food packaged to be consumed away from its location of sale. an establishment serving takeaway. the action or an act of taking out. something taken out or prepared to be taken out.

10. Cost of a monthly gym membership

The cost of membership to attend and utilised a gym. Gym membership is generally paid monthly, or an annual charge based on the gym or fitness studio terms and conditions of membership.

11. *Country name*

     The names of the countries in Asia & Europe region.

12. *Population 2020*

     Population often refers to the number of people in a specific place, whether it be a city or town, region, country, continent, or the planet. Governments often quantify the number of the resident population within their jurisdiction via a census, a process of collecting, evaluating, assembling, and releasing statistics regarding a population.

     So, here is the data for population in 2020 of the countries in Asia & Europe region.

13. *Population 2019*

      So, here is the data for population in 2019 of the countries in Asia & Europe region.

14. *COVID-19 deaths in 2020*

     Number of deaths per 100,000 population in 2020 in Asia & Europe region

15. *Median age*

     Median age is the age that divides a population into two numerically equally sized groups; that is, half the people are younger than this age and half are older. It is a single index that summarizes the age distribution of a population.

16. *Island*

     Number of islands in the Asia & Europe region.

17. *Index of exposure to COVID-19 infections*
18. *Log of average distance to SARS countries*
19. *Female head of government*
20. *Index of institutional trust*
21. *All-cause death count, 2017*
22. *All-cause death count, 2018*
23. *All-cause death count, 2019*
24. *All-cause death count, 2020*
25. *Excess deaths in 2020*

     And many more.

     To visualization the data is used the correlation was plot by seaborn library. The correlation was use compare between each value of column with others. For example, the Happiness Score compared with Family, in this case we should the value between impact the happiness score against the family. The other visualization, we used bar chart to bring out the distribution of all numerical data. Hence, we can analyze each part of columns. For the last case, we describe each column and visualize using table, vertical bar plot, and map plot.

Result

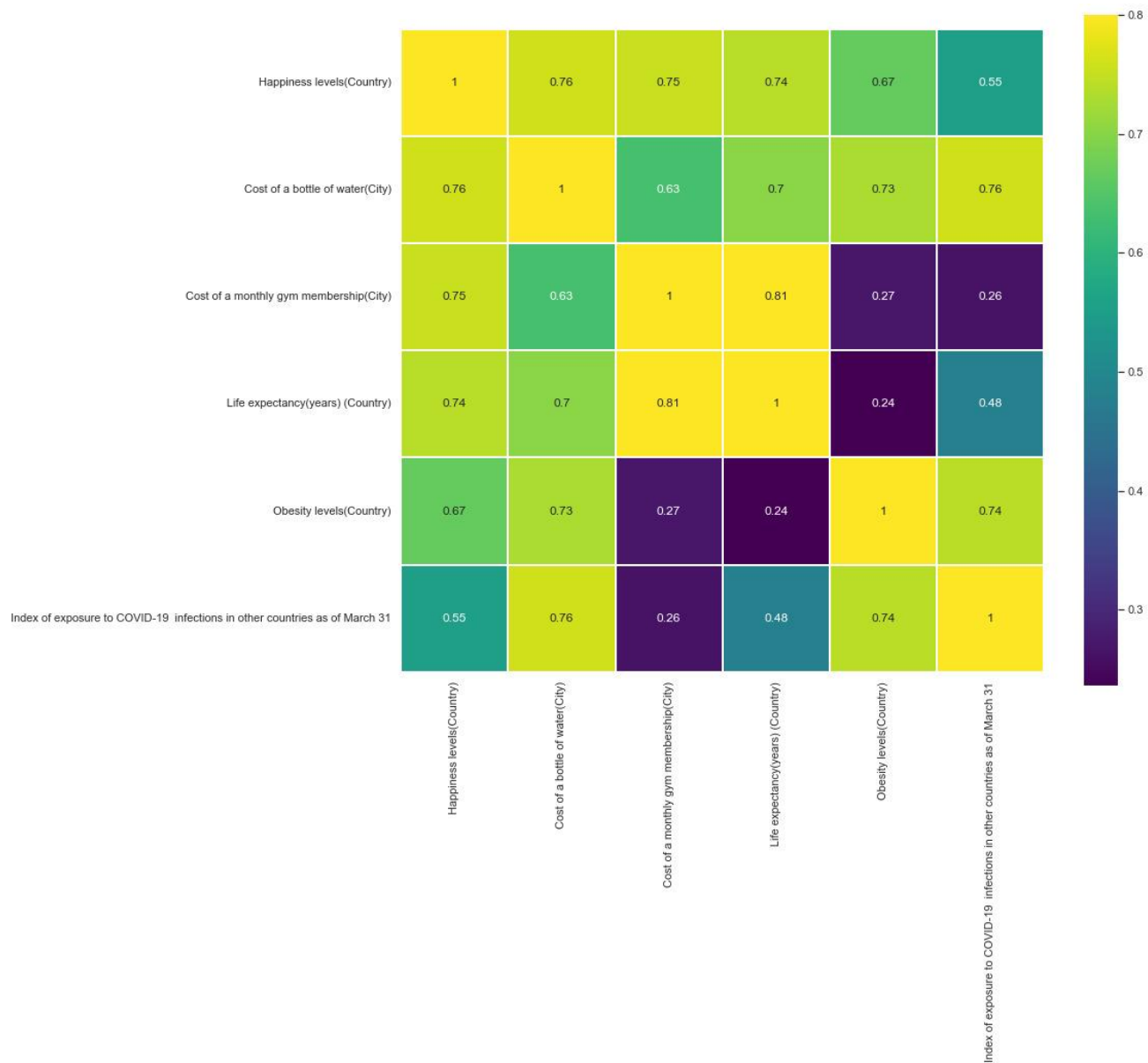- **Co-relations of Happiness levels in Asia**

Fig.3 Correlation Data Value in Heatmap of Asia

Happiness levels(Country)	1.000000
Excess deaths in 2020	0.905885
Cost of a bottle of water(City)	0.759284
Cost of a monthly gym membership (City)	0.752446
Life expectancy(years) (Country)	0.738621
Obesity levels (Country)	0.667471

Index of exposure to COVID-19 infections	0.551458
Annual avg. hours worked	0.433116
COVID-19 deaths in 2020	0.432106
Log of average distance to SARS countries	0.350809
Median age	0.289048
Sunshine hours(City)	0.277648
Island	0.132757
Number of take-out places(City)	0.085007
Outdoor activities(City)	0.072923
WHO Western Pacific Region	0.004166
Female head of government	-0.143617
All-cause death count, 2020	-0.207594
All-cause death count, 2019	-0.209045
All-cause death count, 2017	-0.209322
All-cause death count, 2018	-0.209889
Gini coefficient of income	-0.323705
Index of institutional trust	-0.500690
Pollution(Index score) (City)	-0.581537
Population 2020	-0.620801
Population 2019	-0.627062

- **Co-relations of Happiness levels in Europe**

Fig.3 Correlation Data Value in Heatmap of Europe

Happiness levels(Country)	1.000000
Cost of a bottle of water(City)	0.799724
Index of institutional trust	0.706938
Life expectancy(years) (Country)	0.641579
Median age	0.421156
Cost of a monthly gym membership(City)	0.403203
Female head of government	0.377792
Index of exposure to COVID-19	0.364232
Log of average distance to SARS countries	0.145967

Island	0.107869
COVID-19 deaths in 2020	-0.019498
Number of take-out places(City)	-0.319778
Outdoor activities(City)	-0.473685
Sunshine hours(City)	-0.532071
Obesity levels(Country)	-0.625774
Population 2019	-0.761443
Population 2020	-0.762944
Annual avg. hours worked	-0.769925
Excess deaths in 2020	-0.773395
All-cause death count, 2018	-0.782822
All-cause death count, 2019	-0.783288
Pollution(Index score) (City)	-0.796662
All-cause death count, 2017	-0.797987
All-cause death count, 2020	-0.800694
Gini coefficient of income	-0.807304

The study has been done with several elements present in the dataset, representing the happiness and lifestyle of the citizens of Asia and Europe. The study has found the highest impact on the happiness of the following factors. These are index Excess deaths in 2020 per 100,000 population-relative to the 2017-2019 average, cost of a bottle of water(City), Cost of a monthly gym membership (City), Life expectancy(years) (Country), Obesity levels (Country). These factors have above six positive correlations with the happiness of the citizens. Europe got the highest impact on the Cost of a bottle of water(City), Index of institutional trust, and Life expectancy(years) (Country). This type of factor is relevant for helping to interpret the happiness levels of different regions, especially in Asia & Europe.

Conclusion

Several variables from the dataset, indicating the happiness and way of life of Asian and European residents, were used in the study. According to the study, the following elements have the greatest impact on happiness. Cost of a bottle of water (City), Cost of a monthly gym membership (City), Life expectancy (years) (Country), and Obesity levels are included (Country). These criteria have more than six positive connections with citizen happiness. Europe had the greatest effect on the Cost of a bottle of water (City), Institutional trust index, and Life expectancy (years) (Country). This component is useful for interpreting the happiness levels of various regions, particularly in Asia and Europe. In the future, this study shall be further investigated on an improved dataset with statistical and artificial intelligence-based

models.

References

Riyantoko, P. A. (2020). Southeast Asia Happiness Report in 2020 Using Exploratory Data Analysis. *Ijconsist Journals*, *2*(1), 16-21.

Kaggle data source:https://www.lenstore.co.uk/research/healthy-lifestyle-report/

Nanath, K., Balasubramanian, S., Shukla, V., Islam, N., & Kaitheri, S. (2022). Developing a mental health index using a machine learning approach: Assessing the impact of mobility and lockdown during the COVID-19 pandemic. *Technological forecasting and social change*, *178*, 121560.