

Cross-Country Analysis on Connection Between Financial Lifestyle and Happiness

Anastasiia Dziuba¹, Uroš Sergaš¹

¹University of Primorska, Koper, Slovenia

Abstract

This research aims to investigate if happiness depends on the lifestyle inhabitants can allow themselves from a financial point of view. Furthermore, this paper examines in which countries and regions it is possible to have middle-class expenses with average salary. In addition, a correlation between other features and happiness is observed. For specific correlations, linear models and clustering techniques are presented to identify associations. Obtained results and patterns behind them are further discussed which provides a clearer understanding of the research topic.

Keywords

happiness, cost of living, financial condition, monthly expenses

1. Introduction

Nowadays term “happiness” is getting more widely used. There are different assumptions regarding the reason behind being happy. For instance, Brooks provides different scientific claims regarding the “ingredients of happiness”. Firstly, happiness is described as subjective well-being, which is comprised of genes, circumstances, and habits. Habits, according to the author, are comprised of faith, family, friends, and work. Brooks believes that work in this equation means that the individual can earn success and serve others. However, he claims, that: “Happiness cannot buy satisfaction.” Since people can never have enough money, their wants rise with a paycheck, and the picture of a perfect life is adjusted very quickly [1].

Kahneman argued in his study from 2010, that income influences life satisfaction, but not happiness [2]. However, in 2023, Killingsworth, Kahneman, and Mellers in their article “Income and emotional well-being: conflict resolved”, investigated the relationship between level of income and happiness. For this purpose, a survey was conducted, in which participants were employed adults in the United States, whose yearly income exceeded \$10,000. Using quantile regression to analyze trends in lower and higher income ranges (below and above \$100,000, respectively), the authors concluded that higher income positively influences respondents’ happiness scores. In addition, the authors mentioned that the bottom of happiness distribution rises more rapidly than the top in that range of income. In particular individuals with an annual income below \$100,000 experience a 15% faster rise in happiness compared to those earning more. [3]

Nevertheless, the role of money is considerably rising in society. We can notice that now there is a trend of “buying happiness”. This century can be described as consumerism and of course, it makes an imprint on people’s understanding of happiness. Social media influences it as well, we try to compare our lifestyle with others, it is starting from a small age. Before people did not have that many types of entertainment and could value small things. One of the reasons behind this could be that we are in “endless pursuit of wanting more”[4].


This research aims to observe the relationship between the financial well-being of residents in different countries and their perceived happiness. To achieve this objective various statistical methods will be employed, including heat-map visualisations, correlation analysis, clustering, t-tests, linear and

HCI SI 2023: Human-Computer Interaction Slovenia 2023, January 26, 2024, Maribor, Slovenia

✉ 89232017@student.upr.si (A. Dziuba); uros.sergas@upr.si (U. Sergaš)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

logistic regressions. Nowadays we can find various indexes of evaluating the life on internet. These happiness scores are based on different features. In this research, we will consider the index provided by the World Happiness Report [5].

2. Methodology and gathering data

To study inhabitants' happiness score and their financial condition, exploratory data analysis was applied. For this purpose, the quantitative data was used. The data for this research was obtained from several sources since a dataset containing all required features to perform the desired analysis was not available publicly. Therefore, WHO[6], Kaggle [7], Gapminder [8] were used. Kaggle is a site meant for machine learning practitioners and data scientists, Gapminder is a site displaying time series of development statistics for all countries, WHO stands for World Health Organization, a specialized agency of the United Nations responsible for international public health.

For this research, we used the following region classification, which includes Australia and Oceania, North America, South America, Central America and Caribbean, Europe, East and Southeast Asia, Central Asia, South Asia, Middle East, and Africa.

To omit bias while using the data from different resources, only the common values were used for observation. More details regarding data cleaning will be provided further for the specific datasets.

Every used dataset was available with a .csv extension, was published to the public domain, and is free to use.

2.1. Cost of living dataset

The dataset was provided on Kaggle[7] by Miguel Piedade, a member and contributor of the platform. Miguel Piedade gathered the data from the Numbeo website [9] by web scraping. Numbeo is the world's cost of living database, where the data is updated quarterly. The dataset published on Kaggle was based on prices in 3rd quarter of 2022. The dataset comprised of 4874 cities from different countries, due to data quality, this list was reduced to 760 cities. However, we used only capital cities for observation, which ended up with 133 different countries. The data provided on Kaggle comprised 54 price metrics, in this research we used the following metrics as average required expenses for living: internet, fitness club (monthly fee), monthly pass (local transport), basic (electricity, heating, cooling, water, garbage), 550 min of prepaid mobile tariff local, 4 meals in an inexpensive restaurant, 8 bottles of water, 24 cups of coffee, 2 pieces of clothes. In addition, the expenses for food were based on the minimum basket list for consuming 2400 calories daily and using the Western diet type, which can be found here [10]. Firstly, we included "rent expenses" as well, however, it led to biased results. Since a lot of people live not alone, it is difficult to evaluate the rent price for individuals. We should also consider that some part of the population lives in their properties, therefore, they do not have monthly rent expenses. Overall, we tried to follow the rule of 50/30/20 for the spending division, where 50% of salary goes for our needs, 30% for our wants and 20% supposed to be the savings [11].

2.2. Happiness score dataset

The dataset was published as an appendix to figures for World Happiness Report 2023 [5], which is a publication of the Sustainable Development Solutions Network. The data for the report is powered by Gallup World Poll, a global analytics and advice company [8]. The company collects data using telephone surveys in the majority of cases, however, in some parts of the world (Latin America, Asia, Africa, etc.) face-to-face interviews in randomly selected households. The happiness rankings are based on the respondents' assessment of their lives, in particular a single-item Cantril ladder life-evaluation question (from 1 to 10, where 10 is the highest satisfaction with life). Nevertheless, the data was published in the report for 2023, the survey was conducted in 2022.

2.3. Other datasets

To see if other features impact the life evaluation of people worldwide, we used the following datasets: Life Expectancy, Human Development Index, Coverage Index for essential health services (UHC index), number of suicides, which was further adjusted to suicide rate (number of suicides per capita), child mortality rate, GDP per capita and Net migration. Most of the data was released in 2021, however, the datasets for child mortality rate and number of suicides were last updated in 2019. We obtained the datasets for Life expectancy, UHC index, and Child mortality rate from the World Health Organization [6]. Net migration rate was gathered from Central Intelligence Agency [12]. GDP per capita was included in the dataset published as an annex for World Happiness Report[5]. HDI (Human Development Index) was published in Human Development Reports [13].

A final set of features is provided below in Table 1.

Table 1: Final dataset with features abbreviations and description used for analysis

Column name	Measurement	Variable type	Description
country	categorical	nominal	Name of the country
region	categorical	nominal	Region to which country belongs
savings	numerical	ratio	Average salary after deducting minimal expenses and taxes (in US dollars)
happiness	categorical	ordinal	Metric measured in 2022 by asking participants about their life satisfaction (the scale from 1 to 10)
life	numerical	ratio	Life Expectancy, estimate of average age of participants
hdi	categorical	ordinal	Human Development Index, summary measure of average achievement in key areas of human development
hcov	categorical	ordinal	Coverage index for essential health services
suicide	categorical	ordinal	Suicide rate, number of suicides per 100000 people
child	categorical	ordinal	Child mortality rate, probability that a child will die in the period between birth and 5 years
gdp	numerical	ratio	GDP per capita (in US Dollars)
mig	numerical	ratio	Net migration, difference between immigration and emigration in the particular country

3. Statistical analysis and its results

3.1. Heatmaps

Firstly, let us observe the situation regarding cost of living in the world. Using heatmaps, it is clearly seen in which countries inhabitants are left with negative balance after average monthly expenses (from red to light red) and where people have the opportunity to save some money (shades of blue). However, it can be complicated to look at the specific country, therefore, plots of all world parts are provided below. Figure 1.

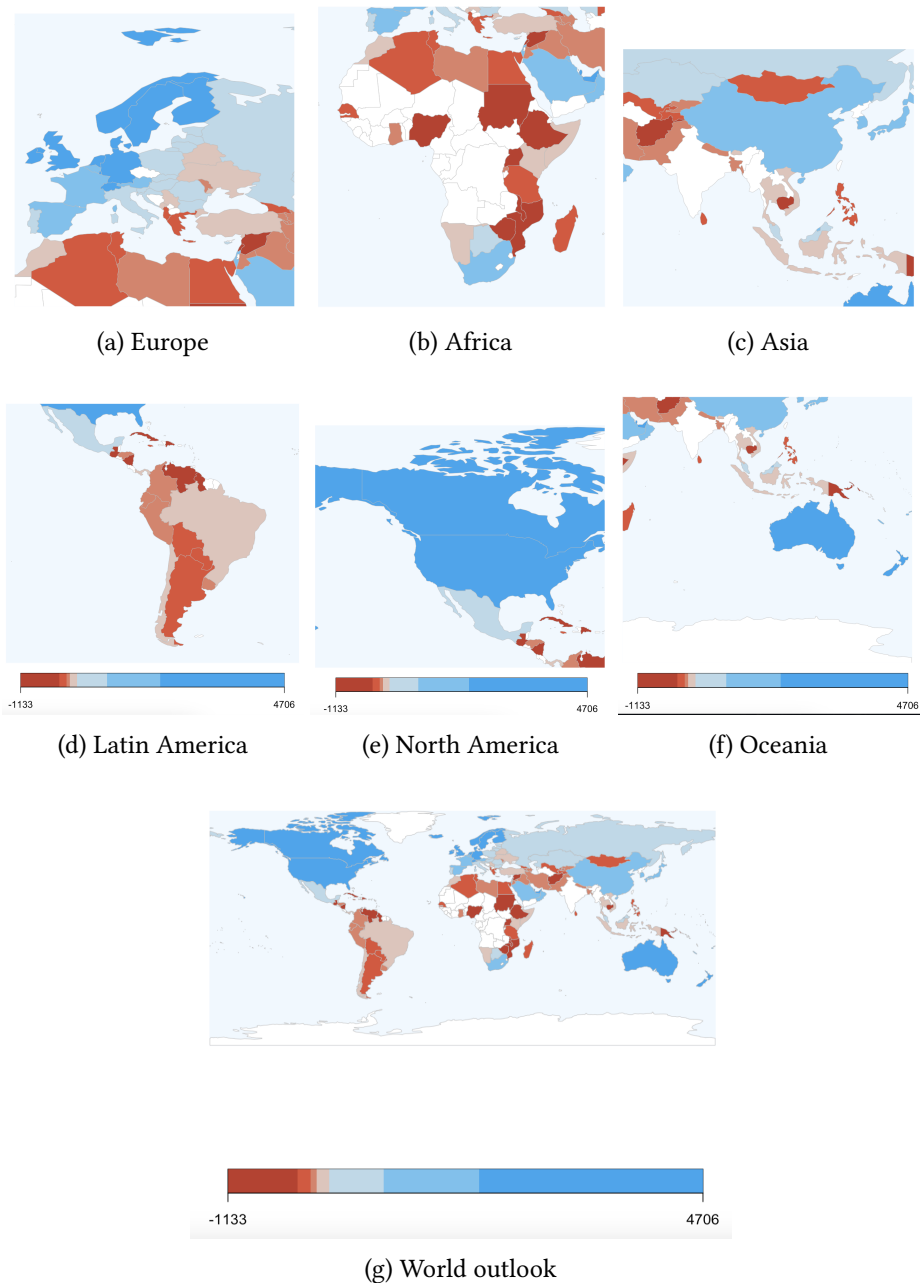


Figure 1: Account balance at the end of the month, US Dollars

3.2. Regions' outlook

Figure 2 shows us the distributional characteristics of groups of savings based on the region classification. We can notice different patterns in the world regions based on the shapes and positions of boxplots. Boxplot for Africa is comparatively short, which suggests that country values there, are low dispersed and normally distributed. On average people there cannot afford all expenses. Australia and Oceania show normal distribution and low dispersion as well, however, the average there is a surplus of 3000\$. Boxplot for East and Southeast Asia suggests a positively skewed distribution of values, therefore there is higher dispersion in the upper quartile. Nevertheless, on average people in this region are left with almost nothing at the end of the month (median ≈ 0). Let us look at the European region, where the values are also highly dispersed and right skewed distribution can be observed. We can suggest that the reason for this lies in different parts of the European region, since in the countries classification we used for research there was no division between countries. There is less dispersion in the 1st and 2nd

quartiles of the boxplot, it is interesting that the maximum values go beyond 4000\$ (the 4th quartile shows high dispersion of values), however, the median is around 750\$. Middle Eastern countries share a similar distribution pattern as European ones, although, the account balance varies there from negative to 1500\$. We can notice high dispersion in the 3rd quartile and a significantly low one in the 2nd quartile. Despite a considerably high maximum, on average people in Middle Eastern countries do not have money for savings at the end of the month. Distribution in the Northern American region tends to be normal, therefore, people there, are left with 2000\$ in their bank account. South Asian and South American countries share similar patterns; however, we can notice a negative outlier in South Asia. People in both regions do not have enough money to afford the average level of life. Unfortunately, the outlook of the Central Asian region is quite biased, only 2 countries are present in this group due to the low data quality.

Overall, we can observe that the financial situation is quite different in world regions, which demonstrates inequality between countries with developing and developed economies based on the United Nations M49 standard [14]. Most of the developed economies are represented by the countries in Europe, North America and Oceania.

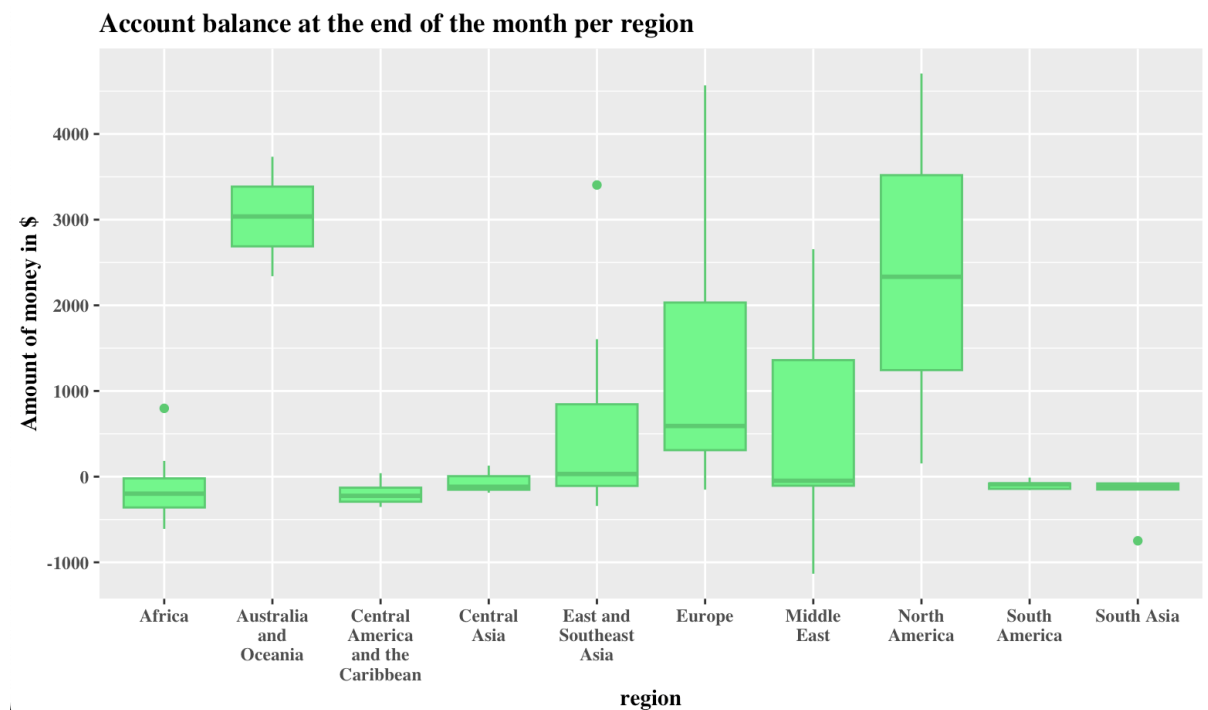


Figure 2: Money left at the end of the month by region, US Dollars

3.3. Correlation

The correlation plot allows us to quantify the direction and strength of the relationship between the included in our research variables. Figure 3 gives us the outlook of correlation in our dataset. We can notice that there is a high positive correlation between happiness ranking and health coverage, human development index, life expectancy, gdp, and financial condition of people. In addition, there high negative correlation between happiness score and child mortality. However, from the plot, we assume that there is a low negative correlation between the suicide rate and other presented variables.

On the other hand, relationships between financial condition and other variables show similar picture as happiness score. The difference is noticeable between life expectancy and financial condition, where the correlation is moderately positive. Finally, we observe a low negative correlation between financial condition and child mortality. It is important to mention that the correlation between inhabitants' bank

accounts and happiness scores is 0.8 in European countries, which is higher than in the world in general. We can suggest that even though in Europe most of the countries are left with a positive balance at the end of the month, people feel happier with higher income. However, the correlation between GDP and happiness rating stays at the same level as in other parts of the world.

The independent t-test was performed to compare the adjusted means of people’s savings in countries with developing and developed economies [14]. There was a significant difference between means of two variables groups ($t(54)=5.076, p < 0.001$). The 95% confidence interval (CI) of the mean difference was from 490\$ (for the lower bound value) to 1130\$. In addition, we observed a significant difference between happiness scores in the examined groups ($t(94)=8.25, p < 0.001$).

Overall, we assume that inhabitants in happier countries tend to live longer, are provided with better health services, and can allow more themselves in financial terms.

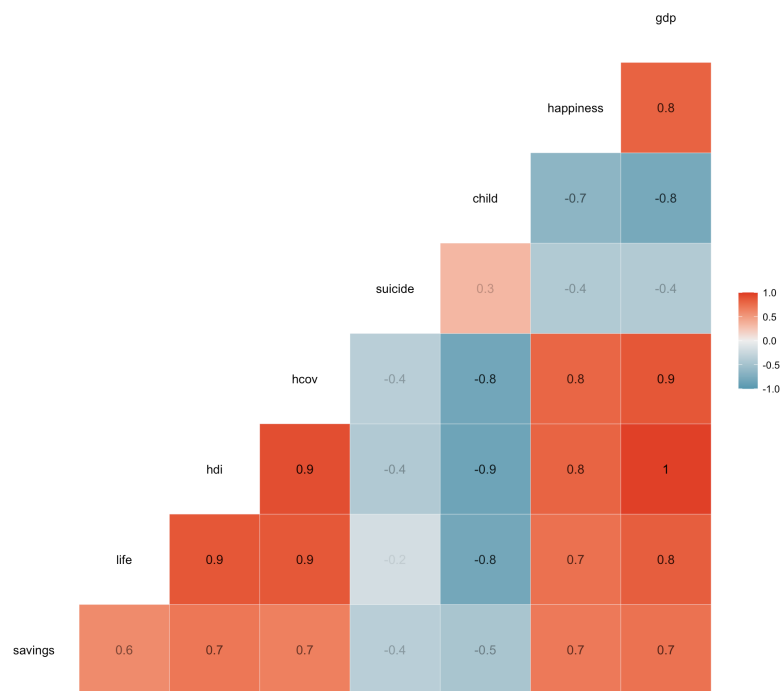


Figure 3: Correlation of used features

3.4. Clustering

This subsection aims to find similar patterns in the relationship between happiness scores and the financial condition of people in different world regions. For this purpose, a k-means clustering algorithm was used, which is based on computing the distance between points in the dataset and a centroid and assigning it to the cluster. Besides world region clustering (Figure 4a), European countries clustering is as well presented below (Figure 4b).

For region clustering the optimal number of clusters according to analysis of all indices except GAP, Gamma, Gplus, and Tau, was 3. Therefore, we can observe that two regions, Australia with Oceania and North America have the highest values of happiness score and potential savings of people. In the middle group, we can find countries from Europe, East and Southeast Asia, and Middle East regions. It is worth mentioning that the leaders of the happiest countries list are European countries, however, considering also financial condition of people living there, we can notice that North America and Pacific countries are leading.

The last group is represented by South and Central Asia, Africa, South America, Central America, and

the Caribbean. The last three regions mentioned are almost on the same level with East and Southeast Asia by happiness ranking, however, they are showing low results in financial condition.

Let us also look at the clustering plot of European countries 4b. Here we can observe 5 clusters, however, one of them is represented only by Switzerland. Finland, which is considered the happiest country in the world [5], can be found in the third cluster, where countries' inhabitants have lower savings than countries in the first cluster (Luxembourg, United Kingdom, Netherlands, Denmark, Norway, and Iceland).

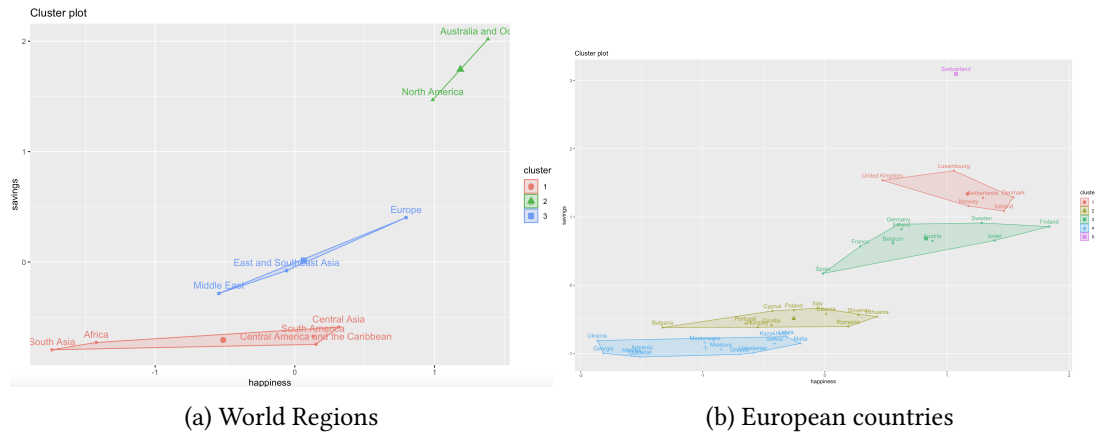


Figure 4: K-Means Clustering based on happiness rate and savings

3.5. Linear regression model

The results of the following linear regression model aim to predict happiness scores based on the amount of money people are left at the end of the month (Figure 5). Looking at the summary closer, we can notice that our model is predicting evenly at both the high and low ends of our dataset since the median value is 0,2. Coefficients allow us to construct the equation for our model. $Y = 0.0006680 * x + 5.365$. As a baseline, if the money at the end of the month equals 0, the happiness score would be 5.365. Then for each 1000\$ of saved money, the happiness score of the country would increase by 0.67 points. In our case, the standard error value (0.00007128) is larger than the coefficient, which shows that the coefficient will most likely not be 0. Therefore, the savings coefficient is 9.37 standard errors away from zero, which is quite far to say that the coefficient will not be 0. The P-value is much smaller than 0.05 ($<2e-16$ and $3.34e-15$ in our case), which means that our coefficient is significant for our model. According to residual standard error, our actual happiness scores are 0.83 points away from predicted ones on average, which is quite high taking into consideration our actual highest scores (7.8). Savings at the end of the month explain 47.77% of the variation with happiness score. Looking at the F-Statistic and p-value we can conclude that the null hypothesis is rejected and there is a relationship between happiness score and the amount of money people are left at the end of the month with average expenses.

Now we can look at the provided graphs (Figure 5). Looking at the Normal Q-Q plot we can suggest that the distribution is negatively skewed, which means that our model possibly overestimates happiness score for countries with smaller income and underestimates for countries with higher one. Scale-Location plot shows us that homoscedasticity is likely to be satisfied for our model. To be sure that homoscedasticity is met, we can perform the Breusch-Pagan test, in our case p-value is 0,175 which is higher than 0,05. Therefore, it indicates that we do not have sufficient evidence to claim that heteroscedasticity is present. Based on the Residuals vs Leverage plot, we notice that there are no influential points in our model (all the points stick very closely to Cook's distance). However, we can remove those points since the model fits the data.

To conclude, taking into consideration all results provided by the linear regression model, there is a relationship between the happiness score of a specific country and the amount of money people are left

at the end of the month.

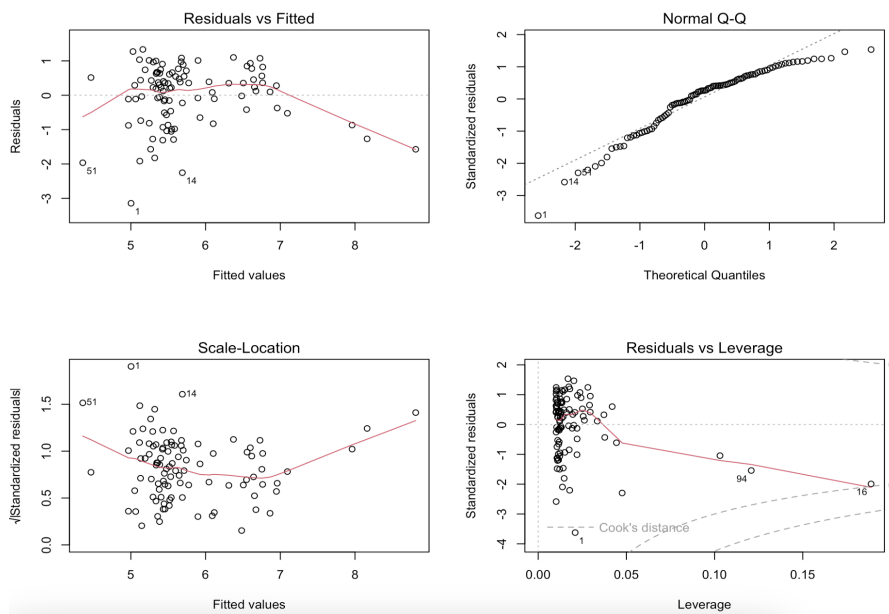


Figure 5: Linear regression model

3.6. Logistic regression model

To better understand the relationship between happiness rate and citizens' financial condition around the world the binary logistic regression was applied. Here we aim to predict if people can cover minimal expenses during the month based on the happiness rate. For this purpose, the threshold of 0 was used for "Savings" (The dataset was randomly divided into trainset and test set in proportion 0.7 and 0.3 respectively).

Based on the obtained results we claim that one unit change in happiness rate increases the log odds of people living in country X to have positive balance accounts after necessary expenses by 1,76. The P-value is greater than the significance level, which indicates that the happiness rate is statistically significant.

In addition, binomial logistic regression was used for observing the relationship between happiness rate and citizens' financial condition around the world. Therefore, we used the happiness score mean as a threshold, which in our case was 5.76. In our original dataset, 43% of sampled countries belong to the group with a happiness rate lower than the mean. The modeling dataset was further randomly divided into a training set, and one used for testing the produced model. Different proportions of splitting data were considered to achieve accurate results. Below the results of one split are described, namely 70% for the training set and 30% for testing of the original dataset. According to the obtained p-value, we claim that the savings variable is statistically significant. In our case for every one dollar increase in savings, the log odds of higher happiness rate increases by 0.004. Considering the intercept of the provided model, we assume that the probability of the country where people are left with no money at the end of the month to achieve a happiness rate higher than 5.76 is approximately 49%.

The presented classification model was used to make the predictions on our test data and further evaluated with a confusion matrix. According to statistics of the confusion matrix, there is 89.66% accuracy of the model predictions. The Kappa metric suggests that our model fit is better than random chance (0.79). 0.517 true positive cases were correctly detected by our model compared to 0.55 true positive cases in our dataset. The ROC curve is provided in Figure 6, the curve is close enough to the upper left corner of the graph which indicates the accuracy of our test. The area under ROC curve

(AUC) is 0.95.

Based on the described results, we assume that the model fits our data. The financial condition of people impacts the happiness level of the country.

Other models were considered as well, namely the relationship between life expectancy and financial condition and, the prediction of a balance account at the end of the month based on the happiness rate. However, the mentioned models did not show accurate results and, therefore, are not presented in detail.

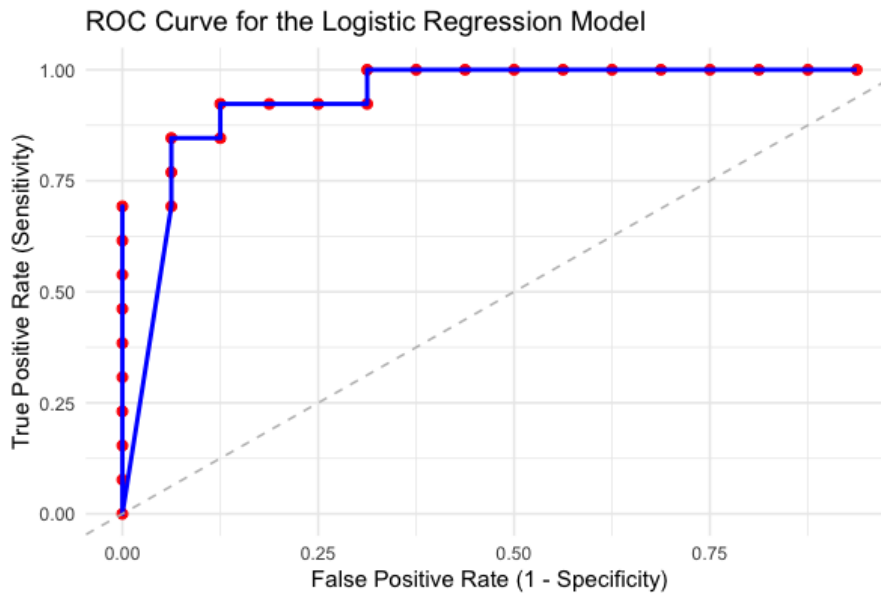


Figure 6: ROC curve for logistic regression model

4. Discussion and Limitations

The applied statistical methods on the used data provide us with a brief overview of the financial condition impact on happiness levels around the world. With other features, such as average life expectancy and human development index we were able to observe how different development levels are interrelated with happiness score and cost of living.

One of the biggest limitations of research is that the used data provides a picture only of capital cities, therefore, we cannot make any conclusions regarding the countries overall. On the other hand, it can be considered more as research on capital cities worldwide. We can suggest that in small cities and villages, the situation with expenses and income is different from capital. Since most of companies vary the prices for products in different countries to ensure a circular flow of economy, but prices inside the country are stable in different cities.

Another limitation would be data quality for the African region, since some data provided for the cost of living is absent, as a result, we cannot make fair conclusions regarding this region. In addition, the population could be divided into groups based on age and marital status, since the expenses in families are built differently from people who live alone. In some families there is only one person who is employed, so in such cases, income is divided between several people.

Since we did not consider the rent expenses in our study, it is difficult to evaluate if the final results are biased. Average property ownership varies from country to country, in some places, it is more affordable to rent the property for living than to own one. We should also consider that not everyone lives in an equal-sized house and there are considerable differences between countries. Property taxation and conditions for purchasing also lie in governmental processes.

We can suggest that people with higher income buy more, therefore, the demand is higher in such countries. For instance, in some developing countries the individuals' picture of prosperity would be totally different from the ones in developed countries. Several decades ago, people did not have such a variety of products and services available, consequently, their demand was lower. Therefore, further research can be done in observing the level of consumerism and happiness scores worldwide.

5. Conclusion

From the presented analysis we can see that there is a relationship between the financial condition of inhabitants and happiness rating in different countries of the world. However, we also observed the limitations that were faced during the research.

Despite the obtained results, it is quite difficult to say that happiness depends on the level of income. Happiness is something abstract, nobody can describe it. There is a picture of happiness in society, where an individual is free, has a lot of money, and can do whatever s/he wants. In real life, our picture of happiness is based on what we consume in terms of information. Information leads our wants, needs, and therefore our expectations of life. Nevertheless, information is free to use in various parts of the world, news feed stays different, and the picture of prosperity, values, satisfaction, happiness, and misery cannot be compared from country to country. Nevertheless, life satisfaction is firstly impacted by the family environment and then by society.

References

- [1] A. C. Brooks, 2020, The 3 equations for a happy life, even during a pandemic, URL: <https://www.theatlantic.com/family/archive/2020/04/how-increase-happiness-according-research/609619/>.
- [2] A. D. Daniel Kahneman, High income improves evaluation of life but not emotional well-being., Proceedings of the National Academy of Sciences of the United States of America (2010). doi:10.1073/pnas.1011492107.
- [3] B. M. Matthew A. Killingsworth, Daniel Kahneman, Income and emotional well-being: A conflict resolved., Proceedings of the National Academy of Sciences of the United States of America (2023). doi:10.1073/pnas.2208661120.
- [4] C. Campbell, The Romantic Ethic and the Spirit of Modern Consumerism: New Extended Edition, Springer Verlag, 2018.
- [5] J. Helliwell, R. Layard, J. D. Sachs, J.-E. D. Neve, S. W. Lara B. Aknin, World Happiness Report 2023 (2023). URL: <https://worldhappiness.report/ed/2023/#appendices-and-data>.
- [6] WHO, 2023, World Health Organization, URL: <https://www.who.int>.
- [7] Kaggle: Your Machine Learning and Data Science Community, 2022, URL: <https://www.kaggle.com/>.
- [8] Gapminder, 2023, URL: <https://www.gapminder.org>.
- [9] Numbeo, 2023, Cost of living, URL: <https://www.numbeo.com/cost-of-living/>.
- [10] Numbeo, 2023, Food Prices, URL: <https://www.numbeo.com/food-prices/>.
- [11] The 50/30/20 rule: how to budget your money more efficiently, 2023. URL: <https://n26.com/en-eu/blog/50-30-20-rule>.
- [12] Central Intelligence Agency, 2023, URL: <https://www.cia.gov>.
- [13] UNDP, 2023, Human Development Reports, URL: <https://hdr.undp.org/data-center/human-development-index#/indicies/HDI>.
- [14] United Nations, 2022, Methodology: Standard country or area codes for statistical use (m49), URL: <https://unstats.un.org/unsd/methodology/m49/>.