Exploring the Impact of Happiness Metrics, Socio-Economic Factors, and Mental Health Disorders on Depression and Quality of Life Across Continents | Aarya Desai | Jeremy Tan | Cassie Kang | Osama Ahmed

Abstract:

In this paper, we analyze two research questions:

RQ1: To what extent do happiness metrics and socio-econmic factors predict depression in continents? RQ2: To what extent do mental health disorders, happiness metrics, and socio-economic factors predict quality of life?

The data used for this research was obtained from the World Happiness Report for information about happiness metrics and from University of Oxford for information about mental health disorders.

For RQ1, our study reveals global patterns in depression predictors. Higher log GDP per capita and freedom to make life choices correlate positively with depression, while greater healthy life expectancy and lower perceptions of corruption are associated with lower depression rates. Generosity, positive affect, low social support, and negative affect impact depression.

For RQ2, our study reveals higher depression and eating disorder percentages predict "Bad Quality of Life," while economic prosperity and freedom enhance the likelihood of "Good Quality of Life." Geographically, the Americas and Oceania are linked to better quality of life, with Asia showing a moderate connection. Surprisingly, alcohol use disorders increase "Good Quality of Life" and drug use disorders also impact it positively. Mental health factors, like schizophrenia and anxiety disorders, also strongly influence "Good Quality of Life." We can attribute this to the complex intricacy of what constitutes a "Good Quality of Life"

Introduction: There is widespread research looking into how happy a country is based on certain socioeconomic and individual-level factors (Ahmadiani, Ferreira, & Kesler 2022). Furthermore, there is also wide-ranging research looking into various mental health disorders and their impact on quality of life (Chigerwe, Holm, Mostret, May, & Bordeaurx 2020). There is a general consensus that the better a countries' socioeconomic status, the higher they score on the happiness index (Ahmadiani, Ferreira, & Kesler 2022). On the other hand, mental health disorders are suggested to be caused by a plethora of factors, such as, genetics, environmental circumstances, financial circumstances, etc. (Evans, Banerjee, Leese, & Huxley 2006). The aim of the current research paper is to gain a better understanding of these two schools of research together, but look at it in a more general context (continent instead of country).

The prevalence of depression is on the rise across the globe (Bell & Blanchflower, 2019). Furthermore, research has indicated that this may have to do with one's socioeconomic status, where individuals with lower socioeconomic statuses are more likely to experience depressive symptoms (Graham & Pinto, 2019). Contrarily, research has also shown that happiness interventions have been shown to reduce depression among individuals (D'raven et a., 2014).

Therefore, there is an established correlation between depression and happiness. The current research paper would like to extend these findings and try to establish a predictive inference of depression based on happiness metrics.

On the other hand, there could be a reverse relationship between these factors as well. To better explain this, it is important to understand the concept of Quality of Life (QoL), which refers to the overall well-being of an individual or population (Teoli & Bhardwaj, 2023). This includes having good personal health (mental, physical, and spiritual), strong relationships, good education and social status, healthy work environment, good income, sense of safety/security, freedom to make choices, and healthy physical environment. Based on this definition, the current study could look into how/to what extent mental health disorders and the other QoL metrics mentioned could predict QoL.

Methods:

Merging Data: As mentioned earlier, two datasets were used for this study: World Happiness Report (2022) and Mental Health Disorders Data (2022). These datasets were merged on common countries and years, and then a larger dataset with 1462 observations was created. It contained 18 variables: `country`, `year`, `life ladder`, `log GDP per capita`, `social support`, `life expectancy at birth`, `freedom`, `perception of corruption`, `generosity`, `positive affect`, `negative affect`, `Alcohol Use Disorders`, `Drug Use Disorders`, `Schizophrenia`, `Bipolar Disorders`, `Anxiety Disorders`, `Eating Disorders`, and `Depression`. We then group our 153 countries into their respecitve continents, as our primary interest lies in discerning variations in our research questions **across different continents**.

Dataset Information: The happiness metrics were a range of factual data collected from census surveys (e.g., Log GDP per capita, life expectancy at birth) and self-reported variables collected from the population of each country in a given year (e.g., life ladder, freedom to make choices, social support, etc). The final variables of this dataset were also self-reported measures of positive and negative emotions which were culminated and calculated into the variables positive affect and negative affect (required to analyze happiness reports for countries).

The mental disorders dataset's variables indicated the percentage of the population that had the stated mental disorder in the given country and year per observation.

Data Cleaning and Manipulation: The merged dataset had multiple observations per country, each observation representing a different year. Though combining information by country was an option, this would involve averaging observations across years per country and removing information about time. However, in relation to happiness metrics and mental health, it is important to maintain information regarding time as research suggests that environmental changes across the years can have an impact on these factors (Arshad, Stevens, & Hide 1993). Since we are interested in how our research questions varies across continents, we then include a categorical variable that represents the continent the country resides in to collapse the countries into different regions (continents).

Another transformation was done in relation to the second research question. Since this question aimed to predict one's lifestyle category based on mental health and happiness metrics, a variable for `lifestyle category` needed to be created. This was done with the help of the variable Life Ladder in the World Happiness dataset. The original variable is called the Cantril ladder, which asks respondents to rate their current lives on a scale of 0 (worst possible life) to 10 (best possible life). In the World Happiness dataset, this variable was an average of all the results collected from citizens of the specific country in the given year (i.e., a float, continuous variable).

To simplify this variable and create an overarching binary variable indicating "Lifestyle Category", i.e., Good Quality of Life (Good QoL) vs. Bad Quality of Life (Bad QoL), the variable life ladder was transformed. This was done using the midpoint of the Life Ladder scale (i.e., 5.5) and then recoding a Lifestyle Category variable to Bad QoL if the observation was below the midpoint and Good QoL if it was above.

EDA:

RQ1 Specific EDA and Steps for Model:

We first identify this is a **prediction problem** and proceed with examining the question as such. Since we are predicting depression % for those in their respective continents, we use linear regression to build a model and select variables based on *priori* selection. We focus on predictor variables: `Log GDP per Capita`, `Social Support`, `Healthy life expectancy at birth`, `Freedom to make like choice`, `Genorisity`, `Perceptions of corruption`, `Positive affect`, `Negative affect`, `Continent`, and `Year` for our outcome variable `Depression %`. These variables were selected after careful examination of the dataset as a whole and incorporating predictor variables that reflect both economic and social conditions of people in their continent and year.

After building the initial benchmark model with these variables, we asses the significance of the variables using p-values, check the multicollinearity using `VIF()`, and check RMSE via cross validation to see if our model will be predict well. If we improve the model, we repeat the same steps as previously stated. To improve our model, we first remove variables to see if there is an impact; in the case, `Social Support`. After seeing no improvement, we examine the `Social Support` variable again and break it up into categories where a continent with high social support (above .7) would be considered "High" and those under would be "Low." We then make an interaction term with `Generosity` and `Social support`. We are further supported by this decison from previous research where `Generosity` and `Social Support` are interconnected(D'Exelle & Riedl 2010). After assessing the model again, we see that `Social Support` and the interaction between `Generosity` and `Social support` becomes significant so we keep this current model.

We then evaluate the influential points to see if there are high leverage points that affect the model. After looking at the plots, we remove observatons 221, which had high leverage. After removing these points, we see both `Genorisity` and `Freedom to make life choices` become more significant so we keep them.

After finalizing our model, we make test data to predict on our model to validate our model.

RQ2 Specific EDA and Steps:

We first identify this is a **prediction problem** and proceed with examining the second questions as such. Since we are predicting QoL for those in their respective continents, we use logistic regression to build a model and select variables based on *priori* selection. Our focus is on predictor variables: `Schizophrenia (%)`, `Bipolar disorder (%)`, `Eating disorders (%)`, `Anxiety disorders (%)`, `Drug use disorders (%)`, `Depression (%)`, `Alcohol use disorders (%)`, `Log GDP per capita`, `Social support`, and `Freedom to make life choices` for our outcome variable, `Quality of Life (%)`. These variables were chosen, taking into account factors that capture both mental health indicators and socio-economic conditions, for individuals in their respective continents and years

After building the initial benchmark model with these variables, we asses the significance of the variables using p-values, check the multicollinearity using `VIF()`and check the accuracy of our model via Cross Validation to see whether our model needs to be improved. If we improve the model, we repeat the same steps as previously stated. We take the same approach as the first research question and transform the social support variable (see ***RQ1 Specific EDA and Steps for Model***). We then plot the variables to see good interaction terms, and settle with `Alcohol use disorders (%)` and `Social Support`. To also add to it, research has suggested that a lack of social support is most likely to increase the risk of substance abuse (Cherry, 2023; Eddie et al., 2019; Horigian et al., 2020). Therefore, it is likely that there is an interaction between the predictors social support and alcohol abuse. For the same, an interaction terms were included in this logistic regression model (i.e., Alcohol use * Social Support) to be in line with current psychological and addictive behaviors literature. After assessing the model again, we see `Alcohol use disorders (%)` become more significant . to * so we keep our current model with the interaction term.

We then evaluate the influential points to see if there are high leverage points that affect our model. After removing observation 381, we see no improvement so we keep our previous model.

We then look at the ROC curve to see the best threshold needed for our model to predict accurately. We then look at the confusion matrix for our final model to asses it with F1, Kappa, and Accuracy.

After finalizing our model, we make test data to predict on our model to validate our model.

Results:

RQ1 Key Figures:

Fig 1. Overview of RQ1 selected Variables
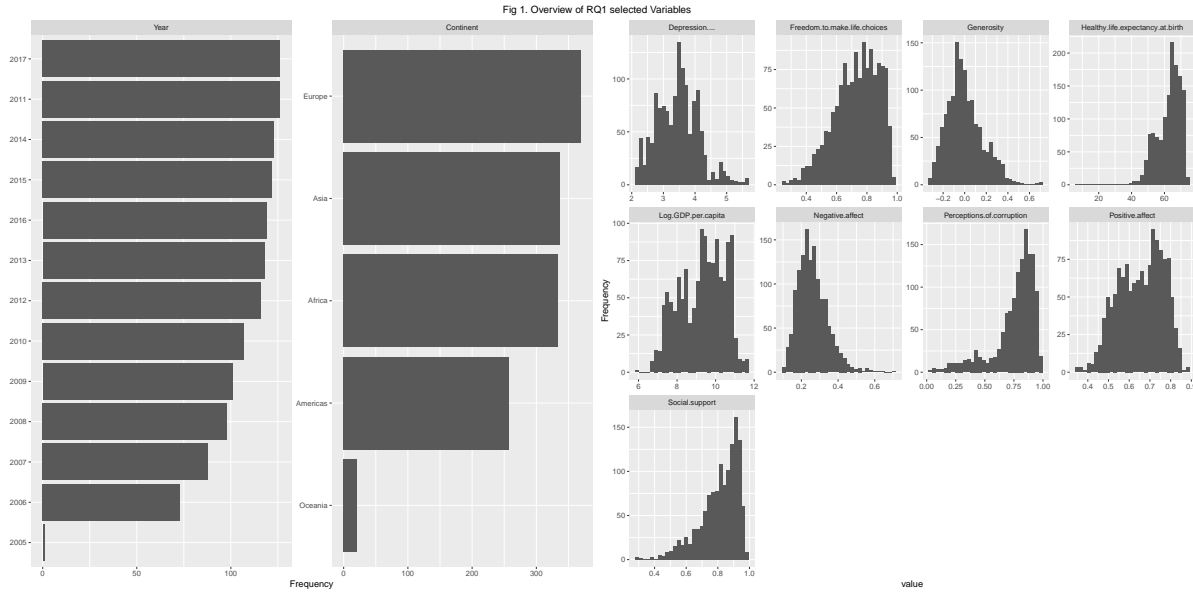
RQ1 Results:

See Table 3 for final model summary.

RQ1 Findings:

1. **Log GDP per capita (0.25):** For each one-unit increase in log GDP per capita, depression tends to rise by 0.25 percentage points.

2. **Healthy life expectancy at birth (-0.02):** A one-unit increase in healthy life expectancy is associated with a decrease of 0.02 percentage points in depression.

3. **Freedom to make life choices (0.35):** When people experience a one-unit increase in the freedom to make life choices, depression tends to increase by 0.35 percentage points.

4. **Perceptions of corruption (-0.63):** A one-unit increase in perceptions of corruption is associated with a decrease of 0.63 percentage points in depression.

5. **Generosity (0.25):** An increase in generosity by one unit corresponds to a rise of 0.25 percentage points in depression.

6. **Positive affect (-0.58):** An increase in positive affect is linked to a decrease of 0.58 percentage points in depression when social support is constant.

7. **Social support category (Low) (0.24):** Compared to the "High" category, being in the "Low" social support category is associated with an increase of 0.24 percentage points in depression.

8. **Negative affect (0.84):** An increase in negative affect by one unit is linked to an increase of 0.84 percentage points in depression.

9. **Interaction: Generosity * Social support category (Low) (1.26):** The interaction effect indicates that the combination of low social support and higher generosity is associated with an increase of 1.26 percentage points in depression. Certainly! Let's interpret the coefficients for the continents:

10. **ContinentAmericas (-0.70):** The coefficient for the Americas continent is -0.70. Being in the Americas is associated with a decrease of 0.70 percentage points in depression.

11. **ContinentAsia (-0.58):** The coefficient for the Asia continent is -0.58. This negative value suggests that, compared to the reference continent, being in Asia is associated with a decrease of 0.58 percentage points in depression.

12. **ContinentEurope (-0.51):** The coefficient for the Europe continent is -0.51. This negative value suggests that, compared to the reference continent, being in Europe is associated with a decrease of 0.51 percentage points in depression.

13. **ContinentOceania (0.01):** The coefficient for the Oceania continent is 0.01. This small positive value suggests that, compared to the reference continent, being in Oceania is associated with a very slight increase of 0.01 percentage points in depression.

Analyzing depression levels across continents provides interesting insights. Higher wealth (Log GDP per capita) surprisingly corresponds to increased depression, challenging expectations. Longer life expectancy is linked to lower depression, while more freedom is unexpectedly tied to higher depression. Lower corruption perceptions are associated with better mental health. Generosity is unexpectedly linked to increased depression. Positive feelings are connected to lower depression, and the importance of supportive communities is clear, with low social support linked to higher depression. Negative feelings predictably contribute to increased depression.

RQ1 Predictions:

1. In a country categorized as part of the Americas continent, during the year 2015, with a log GDP per capita of 9.5, a social support categorized "High", a healthy life expectancy at birth of 70 years, freedom to make life choices of 0.7, generosity of 0.1, and perceptions of corruption at 0.05, the estimated depression percentage is predicted to be around 3.37%.

2. In a country situated in the Asia continent, during the year 2014, with a log GDP per capita of 10.0, a social support categorized as "Low", a healthy life expectancy at birth of 75 years, freedom to make life choices of 0.8, generosity of 0.2, and perceptions of corruption at 0.02, the estimated depression percentage is predicted to be approximately 4.08%.
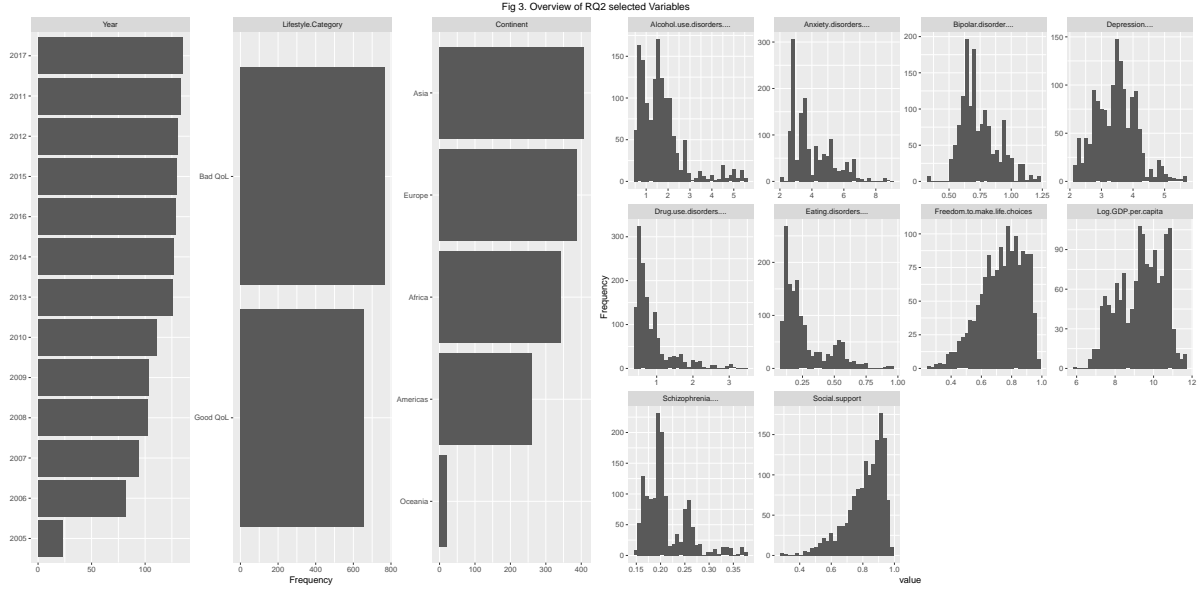
RQ1 Metrics:

Table 1: RQ1 Model Key Metrics

Metric	Value
RMSE	0.5789729
R^2	0.2195582
Adj. R^2	0.2350000

RQ1 Plot:

Fig 2. Impact of Variables on Depression

RQ2 Key Figures:
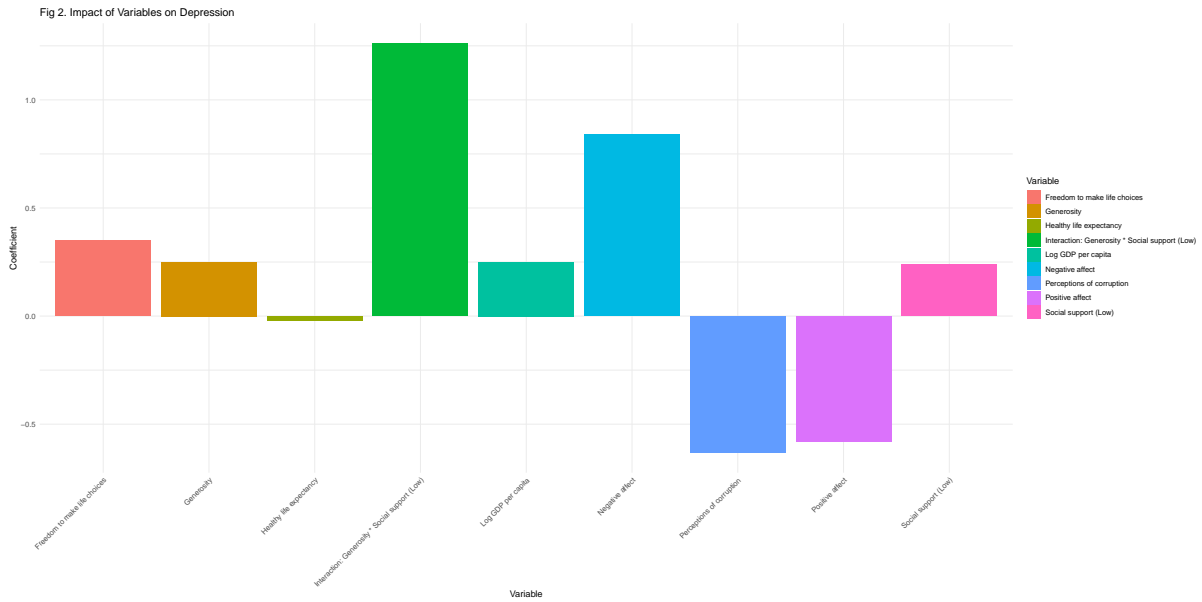
Fig 3. Overview of RQ2 selected Variables

RQ2 Results:

See Table 4 for final model summary.

RQ2 Findings:

1. **Depression (%) (-0.446):** For a one-unit increase in the percentage of depression, the log-odds of being in the specified lifestyle category decrease by 0.446.

2. **Schizophrenia (%) (6.033):** For a one-unit increase in the percentage of schizophrenia, the log-odds increase by 6.033.

3. **Bipolar disorder (%) (0.010):** For a one-unit increase in the percentage of bipolar disorder, the log-odds increase by 0.010.

4. **Eating disorders (%) (5.537):** For a one-unit increase in the percentage of eating disorders, the log-odds increase by 5.537.

5. **Anxiety disorders (%) (-0.307):** For a one-unit increase in the percentage of anxiety disorders, the log-odds decrease by 0.307.

6. **Log GDP per capita (1.823):** For a one-unit increase in the log GDP per capita, the log-odds increase by 1.823.

7. **Freedom to make life choices (6.578):** For a one-unit increase in the freedom to make life choices, the log-odds increase by 6.578.

8. **Social support category (Low) (-0.007):** Being in the low social support category is associated with a log-odds decrease of 0.007.

9. **Alcohol use disorders (%) (0.351):** For a one-unit increase in the percentage of alcohol use disorders, the log-odds increase by 0.351.

10. **Drug use disorders (%) (0.724):** For a one-unit increase in the percentage of drug use disorders, the log-odds increase by 0.724.

11. **Continent (Americas) (3.025):** Being in the Americas continent is associated with a log-odds increase of 3.025.

12. **Continent (Asia) (0.829):** Being in the Asia continent is associated with a log-odds increase of 0.829.

13. **Continent (Europe) (0.478):** Being in the Europe continent is associated with a log-odds increase of 0.478.

14. **Continent (Oceania) (11.829):** Being in the Oceania continent is associated with a log-odds increase of 11.829.

The model provides valuable insights into predicting either "Good Quality of Life (QoL)" or "Bad QoL." Notably, higher percentages of depression and eating disorders are associated with a predicted "Bad QoL," while economic prosperity (measured by log GDP per capita) and the freedom to make life choices contribute to a higher likelihood of "Good QoL." Geographically, being in the Americas or Oceania is linked to a better quality of life, with Asia showing a moderate connection to improved QoL. Surprisingly, alcohol use disorders increase the likelihood of "Good QoL," while drug use disorders also have a substantial impact. Mental health factors, such as schizophrenia and anxiety disorders, strongly influence what is considered a "Good Quality of Life." These findings underscore the complex nature of "Quality of Life," where various factors, including disorders, have a complex relationship with each other.

RQ2 Predictions:

1. In a country categorized as part of the Americas continent, during the year 2015, with a log GDP per capita of 3.0, a social support categorized as "Low," a healthy life expectancy at birth of 70 years, freedom to make life choices of 0.9, perceptions of corruption at 0.01, and the following mental health disorder percentages: Depression: 2.1%, Schizophrenia: 0.12%, Bipolar disorder: 0.1%, Eating disorders: 0.1%, Anxiety disorders: 0.11%, Alcohol use disorders: 0.1%, and Drug use disorders: 0.1%

2. In a country situated in the Asia continent, during the year 2014, with a log GDP per capita of 10.0, a social support categorized as "High," a healthy life expectancy at birth of 75 years, freedom to make life choices of 0.8, perceptions of corruption at 0.02, and the following mental health disorder percentages: Depression: 3.6%, Schizophrenia: 0.4%, Bipolar disorder: 0.5%, Eating disorders: 0.7%, Anxiety disorders: 0.12%, Alcohol use disorders: 0.3%, and Drug use disorders: 6%

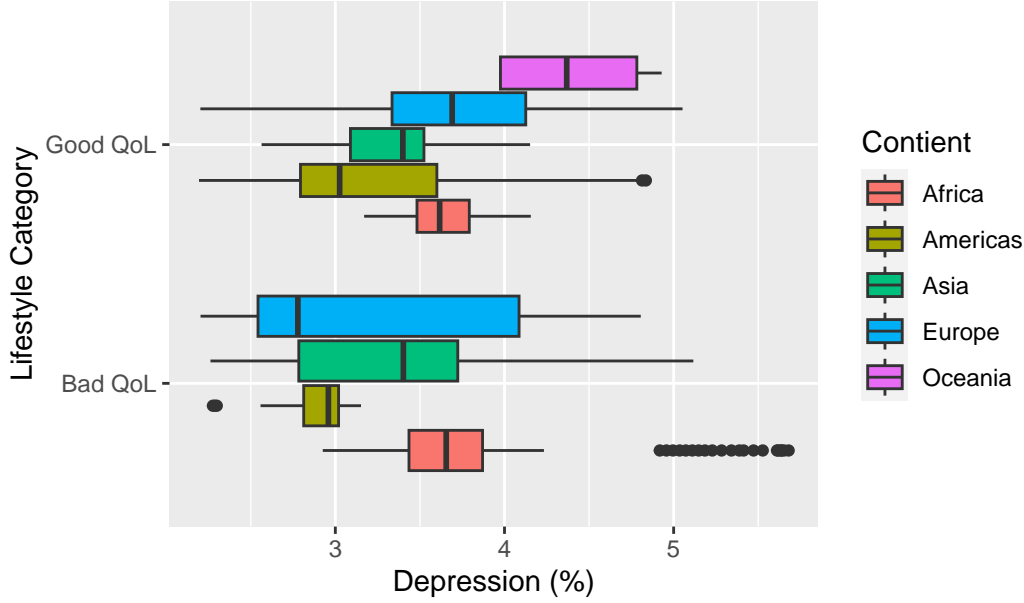Case 1 is predicted to have "Bad QoL" while Case 2 is predicted to have "Good QoL."

RQ2 Metrics:

Table 2: RQ2 Key Metrics

	Accuracy	Kappa	F1
Value	0.8988053	0.7958218	0.8881988

RQ2 Result Plot

Figure 4. Box Plot of Lifestyle Category vs. Depression (%

Conclusion:

When analyzing both models, we are confident that both models can accurately predict both quality of life and depression within different continents after examining RMSE, and a combination of F1, Kappa, and Accuracy for the respective models. We feel, however, that we need more data for some continents (such as Oceania) and more years to further enhance both our models.

References

Bell, D. N., & Blanchflower, D. G. (2019). The well-being of the overemployed and the underemployed and the rise in depression in the UK. *Journal of Economic Behavior & Organization*, *161*, 180-196.

Cherry, K. Mse. (2023, March 3). *A social support system is imperative for Health and well-being.* Verywell Mind. https://www.verywellmind.com/social-support-for-psychological-health-4119970#:~:text=1%20Poor%20social%20support%20has,Alcohol%20use

D'raven, L. T. L., Moliver, N., & Thompson, D. (2015). Happiness intervention decreases pain and depression, boosts happiness among primary care patients. *Primary health care research & development*, *16*(2), 114-126.

Eddie, D., Hoffman, L., Vilsaint, C., Abry, A., Bergman, B., Hoeppner, B., ... & Kelly, J. F. (2019). Lived experience in new models of care for substance use disorder: a systematic review of peer recovery support services and recovery coaching. *Frontiers in psychology*, *10*, 1052.

Graham, C., & Pinto, S. (2019). Using Well-Being Metrics to Assess Social Well-Being and Ill-Being: Lessons from Rising Mortality Rates in the United States. *The Economics of Happiness: How the Easterlin Paradox Transformed Our Understanding of Well-Being and Progress*, 319-353.

Horigian, V. E., Schmidt, R. D., & Feaster, D. J. (2021). Loneliness, mental health, and substance use among US young adults during COVID-19. *Journal of psychoactive drugs*, *53*(1), 1-9.

Teoli D, Bhardwaj A. Quality Of Life. [Updated 2023 Mar 27]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK536962/#

The world by income and region. WDI - The World by Income and Region. (2023). https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html

Xu, X., Geng, F., & Xu, H. (2021). Understanding Loneliness Among Older Adults in Rural China: A Mixed-Methods Study. *Applied Research in Quality of Life.* https://doi.org/10.1007/s10902-021-00478-y

Hanssen, L., Deindl, C., & Lechler, B. (Year not provided). Determinants of Well-Being and Health in Old Age: The Role of Living Arrangements and Partnership Status. *PLOS ONE, 15*(10), e0241302. https://doi.org/10.1371/journal.pone.0241302

Siemsen, E. H., Cappaert, R. A., Gonzalez, J. T., & Wilson, R. R. (2007). The WHOQOL-OLD in the United States: Methodological and Structural Issues. *Quality of Life Research, 16*(1), 1-12. https://doi.org/10.1007/s11136-006-9002-6

Ortega, J., & Hergovich, P. (2011). The Strength of Absent Ties: Social Integration via Online Dating. *SSRN.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1730092

The World by Income and Region. (2023). *WDI - The World by Income and Region.* https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html

Appendix

Table 3.

Predictors	Estimates	std. Error	CI	p
			Depression (%)	
(Intercept)	3.50	0.65	2.23 − 4.77	**<0.001**
Continent [Americas]	-0.69	0.07	-0.82 − -0.56	**<0.001**
Continent [Asia]	-0.57	0.06	-0.70 − -0.45	**<0.001**
Continent [Europe]	-0.49	0.07	-0.64 − -0.35	**<0.001**
Continent [Oceania]	0.03	0.15	-0.26 − 0.32	0.860
Year [2006]	-0.42	0.58	-1.55 − 0.72	0.470
Year [2007]	-0.39	0.58	-1.52 − 0.74	0.497
Year [2008]	-0.31	0.58	-1.44 − 0.83	0.597
Year [2009]	-0.41	0.58	-1.54 − 0.72	0.476
Year [2010]	-0.41	0.58	-1.54 − 0.72	0.478
Year [2011]	-0.40	0.58	-1.53 − 0.73	0.489
Year [2012]	-0.42	0.58	-1.55 − 0.71	0.468
Year [2013]	-0.43	0.58	-1.56 − 0.70	0.456
Year [2014]	-0.48	0.58	-1.61 − 0.65	0.405
Year [2015]	-0.46	0.58	-1.59 − 0.68	0.430
Year [2016]	-0.45	0.58	-1.58 − 0.68	0.434
Year [2017]	-0.47	0.58	-1.60 − 0.66	0.413
Log GDP per capita	0.26	0.03	0.20 − 0.31	**<0.001**
Healthy life expectancy at birth	-0.02	0.00	-0.03 − -0.01	**<0.001**
Freedom to make life choices	0.33	0.16	0.01 − 0.65	**0.041**
Generosity	0.27	0.12	0.03 − 0.52	**0.028**
Social support category [Low]	0.13	0.05	0.03 − 0.23	**0.014**
Perceptions of corruption	-0.59	0.11	-0.80 − -0.38	**<0.001**
Positive affect	-0.47	0.26	-0.98 − 0.03	0.068
Negative affect	0.80	0.23	0.34 − 1.25	**0.001**

	Estimate	Std. Error	z value	Pr(>|z|)
Generosity × Social support category [Low]	-0.19	0.31	-0.80 – 0.41	0.530
Observations	1317			
R^2 / R^2 adjusted	0.249 / 0.235			

Table 4.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-24.815	2.704	-9.178	0.000
'Depression (%)'	-0.446	0.251	-1.775	0.076
'Schizophrenia (%)'	6.033	5.083	1.187	0.235
'Bipolar disorder (%)'	0.010	1.278	0.008	0.994
'Eating disorders (%)'	5.537	1.719	3.220	0.001
'Anxiety disorders (%)'	-0.307	0.151	-2.037	0.042
'Log GDP per capita'	1.823	0.253	7.198	0.000
'Freedom to make life choices'	6.578	0.911	7.224	0.000
'Alcohol use disorders (%)'	0.351	0.142	2.475	0.013
'Social support category'Low	-0.007	1.172	-0.006	0.996
'Drug use disorders (%)'	0.724	0.335	2.160	0.031
ContinentAmericas	3.025	0.579	5.229	0.000
ContinentAsia	0.829	0.449	1.845	0.065
ContinentEurope	0.478	0.549	0.871	0.384
ContinentOceania	11.829	462.833	0.026	0.980
Year2006	-0.769	0.861	-0.893	0.372
Year2007	0.242	0.843	0.287	0.774
Year2008	-0.187	0.852	-0.219	0.826
Year2009	0.159	0.842	0.189	0.850
Year2010	0.759	0.840	0.903	0.366
Year2011	-0.226	0.826	-0.274	0.784
Year2012	0.309	0.830	0.372	0.710
Year2013	-0.018	0.826	-0.022	0.982
Year2014	0.306	0.828	0.370	0.711
Year2015	0.354	0.827	0.428	0.669
Year2016	0.115	0.827	0.139	0.890
Year2017	0.439	0.826	0.531	0.595
'Alcohol use disorders (%)':'Social support category'Low	-0.258	0.727	-0.355	0.723