UNIVERSITY OF SOUTH FLORIDA

BUSINESS ANALYTICS AND INFORMATION SYSTEMS

ISM 6137 - STATISTICAL DATA MINING

**LIFE SATISFACTION RATE**

SRI SATYA SAKETH CHIDEPUDI

**Table of Contents**

|  |  |
| --- | --- |
| **Title** | **Page No** |
| Executive Summary | 3 |
| Problem Definition & Significance | 3 |
| Prior Literature | 4 |
| Data Source / Preparation & Variable Choice | 4 |
| Descriptive Analysis & Data Visualizations | 6 |
| Models | 9 |
| Quality Checks | 11 |
| Recommendations | 12 |
| References | 12 |
| Appendix: R Code | 13 |

1. **Executive Summary**

**Problem Statement:**

Lower life satisfaction rate has a negative impact on the government in different ways. Some of them are migration to the other countries, fall in the tourism rate and so, thereby effecting the economy of the country. It is crucial to evaluate the life satisfaction rate and understand the significant factors influencing it, to the betterment of the country.

**Data Analysis:**

Our work focuses on predicting the life satisfaction rate of each country based on the 18 other variables like Crime Rate, Unemployment rate, Pollution, GDP, Health Index, Cost of Living Index, Pollution, Quality of Life, Population Density and Affordability. Our data has 165 rows with some null values which has been handled by using K-means clustering.

**Key Findings:**

The health index is the strongest variable contributing to the life satisfaction of the country followed by GDP and cost of living. It shows that countries that have better health facilities, that are able to create more value through the production of products and services and that have better living conditions record higher life satisfaction rates. And also, better employment opportunities and safe environment for its citizens slightly drive the life satisfaction of the country.

1. **Problem Definition & Significance**

Lower life satisfaction rate has negatively impacted the country in many ways. Now let us dig deeper into how and why these factors are important.

**Tourism:**

The total contribution of Travel & Tourism to GDP was **$1,509.2 billion**. Tourism not only contributes to the economy of the country but also drives better employment opportunities in airlines industry, as tourist guides and in hotel management. And, the countries that have better facilities and better life satisfaction rates are the ones that attract more tourists.

The countries with the life expectancy close to or higher than the average level of 5.4 are the countries with higher tourism. The top 20 tourists’ attractions like Netherlands, United Kingdom, Germany, Canada, United States of America have a life expectancy higher than the average level.

**Migration:**

Migration effects the population count, productivity level and the economy of the home country. Stats show that **more than half of the US startups** that value for **$1 Billion** or more are founded by at least 1 non-native individual. This thereby increases the economy of the host country, reducing the economy of the home country. Individuals prefer to move to countries like U.S., Germany, Canada, Australia to name a few as they provide higher quality of life.

**Well-being of the society:**

Although the above factors directly or indirectly are taken into consideration for the wellbeing of the individuals, let us put some more light into this. Every nation looks for the wellbeing of the individuals and as a result wishes to understand how happy or content are its citizens with the facilities provided and the underlying factors that contribute to a higher quality of life for its citizens.

1. **Prior Literature**

Listing out some prior articles and their key findings:

* Environment and happiness: Valuation of air pollution using life satisfaction data: Heinz Welsch

This paper evaluates the air pollution of a country based on particles like NO2 and Lead. The paper explains that the air quality at a certain place plays a significant role in driving the life satisfaction.

* A Reassessment of the Relationship between GDP and Life Satisfaction: Eugenio Proto, Aldo Rustichini

This paper focuses on the relation between GDP and life satisfaction. It is reported that the curve gradually rises for countries with GDP below 15,000 USD and then tends to flatten. For countries with GDP above 17,000 USD, the probability of reporting the highest level of Life satisfaction changes with a max range of 2%.

* The Effect of Crime on Life Satisfaction: Mark A. Cohen

The author happens to see that the crime related variables explained only a 7% about the Life Satisfaction rate. He found that the victims of crime related accidents report slightly lower levels of life satisfaction and also found evidence that the life satisfaction rate was relatively lower for individuals who lived in high‐crime areas than those who lived in low‐crime countries.

* Health and Social Factors Related to Life Satisfaction: Erdman Palmore, Clark Luikart

Self-rated health was by far the strongest variable related to life satisfaction followed by organizational activity and belief in internal control. Having a confidant, performance status, employment, and social activity are significantly associated with satisfaction among men, but not among women. Income and education were more strongly related to satisfaction among the younger middle-aged and among those with below average incomes. The rest show little to no relation.

1. **Data Source/ Preparation & Variable Choices**

We have collected different variables from multiple sources.

* Life Satisfaction (Dependent variable) - <https://ourworldindata.org/happiness-and-life-satisfaction>
* Crime Rate – <https://worldpopulationreview.com/country-rankings/crime-rate-by-country>
* Unemployment rate – <https://en.wikipedia.org/wiki/List_of_countries_by_unemployment_rate#cite_note-14>
* Pollution - <https://data.worldbank.org/indicator/EN.ATM.PM25.MC.M3>
* GDP - <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?view=chart>
* Cost of Living Plus Rent Index Groceries Index - <https://www.kaggle.com/morriswongch/cost-of-living?select=cost_of_living_numbeo_2020.csv>
* Climate Index - <https://www.kaggle.com/dumbgeek/countries-dataset-2020?select=Quality+of+life+index+by+countries+2020.csv>
* Density pop./mi2 - <https://www.kaggle.com/dumbgeek/countries-dataset-2020?select=Pupulation+density+by+countries.csv>
* Health Index - <https://www.kaggle.com/dumbgeek/countries-dataset-2020?select=Health+care+index+by+countries+2020.csv>

|  |  |
| --- | --- |
| **Variables** | **Measurement & Justification** |
| Life Satisfaction (Dependent variable) | We are predicting the life satisfaction rate of different countries |
| Crime Rate | The lesser the crime rate, the more the life satisfaction of that country |
| Unemployment rate | Ratio of Unemployed population to the total population. The more employment opportunities offered by the country, the more the employment rate and the more content are its citizens |
| Pollution | The higher the presence of PM2.5 level in the environment the lower the life satisfaction rate |
| GDP | Current GDP of each country in US$.  The better the GDP, the better benefits offered to the citizens and the more content are the individuals in that country. |
| Health Index | The healthier the population, the happier and satisfied they are |
| Density pop./mi2 | Number of human inhabitants per square miles.  Life satisfaction of a country can both increase or decrease w.r.t inhabitants per square miles, as lower population means more benefits in terms of health facilities or so, but also lower productivity on the flip side. |
| Cost of Living Plus Rent Index | Cost of living index is relative to USA. Proportion of income w.r.t. the living expenses effects the life satisfaction of the individual which further contributes to that of a country. |
| Climate Index | Climate index is a simple measure of the degree. Being exposed to daily average temperatures above 90 degrees Fahrenheit reduces overall emotional well-being by around 6% of a standard deviation. |
| Affordability Index | Affordability to basic needs has a direct impact on the life satisfaction. Affordability has a positive impact on the satisfaction of the individuals |

**Data Prep/Cleaning**

* Joined multiple datasets using reduce function in tidyverse library (left- join based on countries)
* Converted the response variable, life satisfaction rate into a categorical column for implementing Ordinal Logistic Regression

(i) If value <4 – **Low**

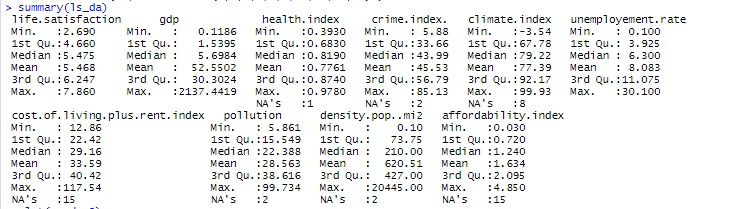
(ii) If value ranges between 4-6 – **Medium**

(iii) Value >6 – **High**

* Imputed the missing values using K-means clustering
* Standardized the numerical features using scale function

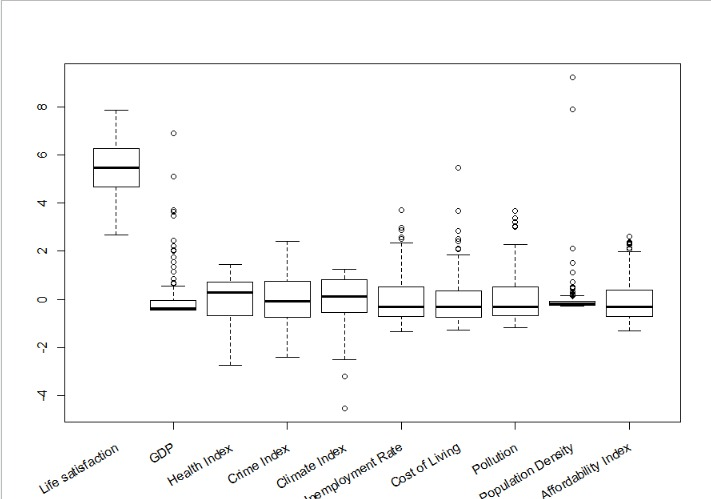
1. **Descriptive Analysis & Data Visualizations.**

**#Descriptive Analysis – Our Data before standardization**



* The minimum life satisfaction rate of 2.69 is observed in Afghanistan and the maximum of 7.86 is observed in Finland with an average of 5.46.
* The GDP value ranges between 0.11\*1010 in Comoros and 2137 \* 1010 in United States in America and an average of 52.55\*1010
* The country with a highest health facility is recorded in Japan (0.97) and the lowest is recorded in Sierra Leone (0.39) and the average health index recorded is 0.77
* Venezuela has the maximum crime rate (85.13) recorded and Comoros has the lowest with a value 5.88
* The country with the most favorable climate of 99.93 is Venezuela and the one with the least is Mongolia (-3.54) and the average is 77.39
* The country with the best employment opportunities to all its citizens is Qatar with a value of 30.100 and the one with the worst is South Africa, a value of 0.10

**#Descriptive Analysis – Our Data after standardization**

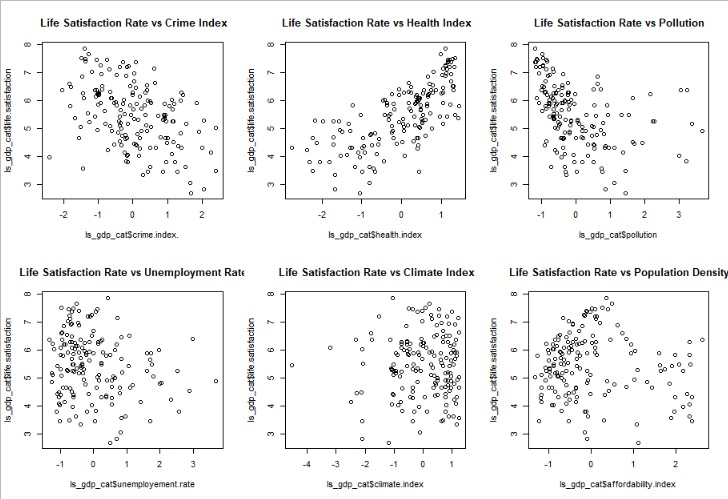


After standardization, we can see that now the mean for all the predictors is 0 with a standard deviation of 1.

From the box plot above, we can say that the variable GDP has the highest outliers, followed by Population density, Cost of Living, Affordability and Unemployment rate.

No outliers for Health index, Crime index and Climate index are observed.

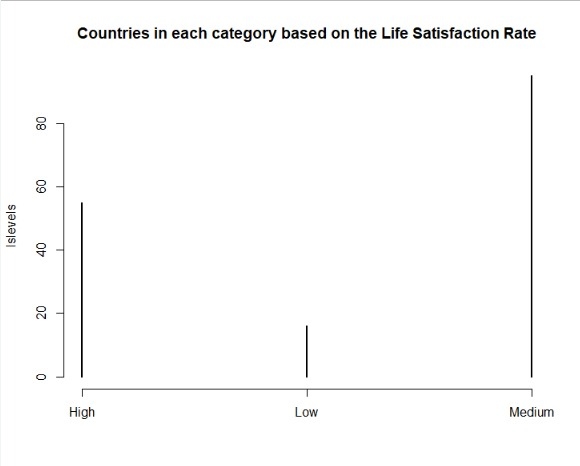
**#Linearity Plot of Life Satisfaction rate with different predictors**



We can see negative slope for crime and pollution and a positive slope for health index.

To some extent we can see negative slope for unemployment and affordability index.

**Classification of our Predicted Variables**



Based on the Life Satisfaction rate, we have classified our data into 3 different categories

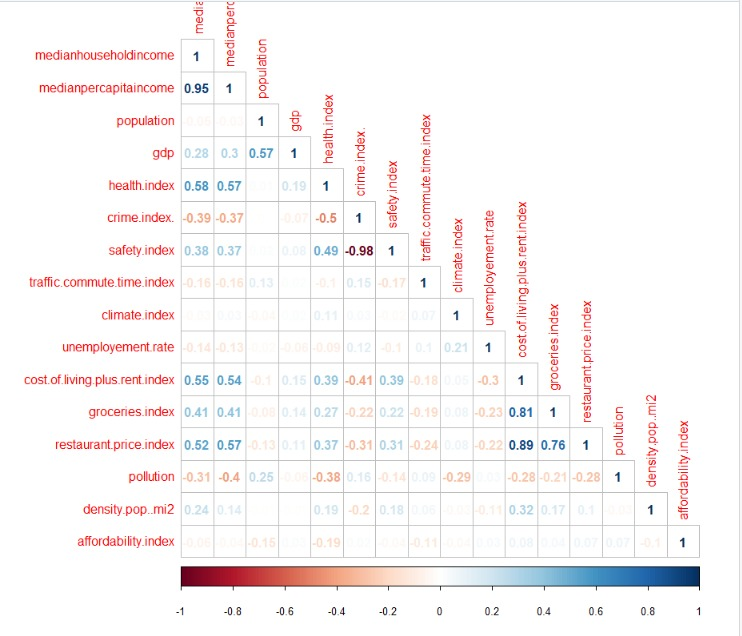
(i) If value <4 – Low

(ii) If value ranges between 4-6 – Medium

(iii) Value >6 – High

Our dataset has 55 countries with higher life satisfaction rate, 16 with lower and 95 countries with medium life satisfaction rate.

**Finding the correlation between the independent variables**

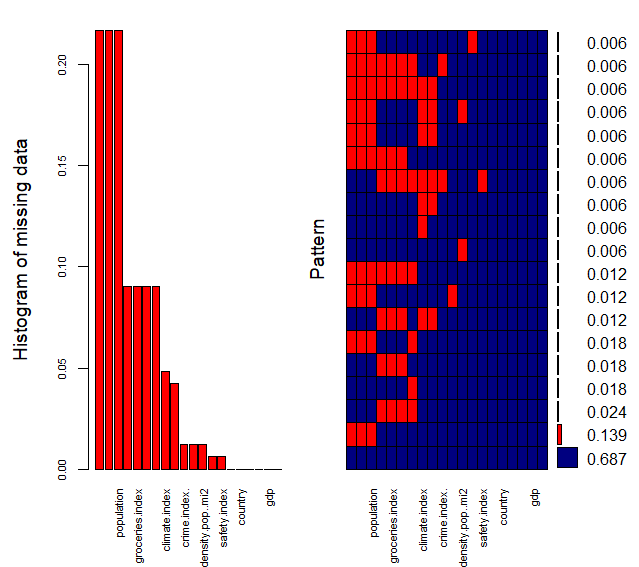


The correlation between median household income & median per capita income seem to be higher i.e. 95%

We also, wee that the correlation between crime index and safety index is as high as 98%, that is why we eliminate safety index from our model.

Also, we have a higher correlation between grocery index, cost of living plus rent index and restaurant price. Therefore, in our model we only use the cost of living plus rent index variable.

**Checking the missing the values in the data-frame**

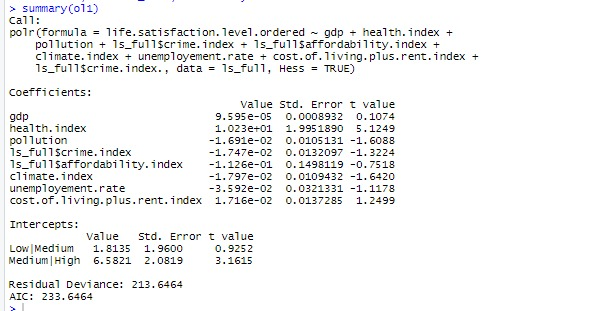


We see that the variables, household income, per capita income and population have the highest missing values. The variable GDP has 0 missing values.

1. **Models**

**Ordinal Logistic Regression**

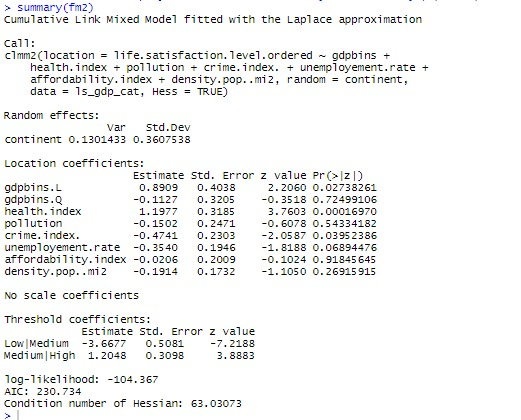
In Ordinal Logistic Regression we need to specify the order for the categories in the response variable.



* The greater the magnitude of T, the greater the evidence against the null hypothesis.
* In this model, we see only the health index value is significant.

**Including Random Effects in Ordinal Regression**

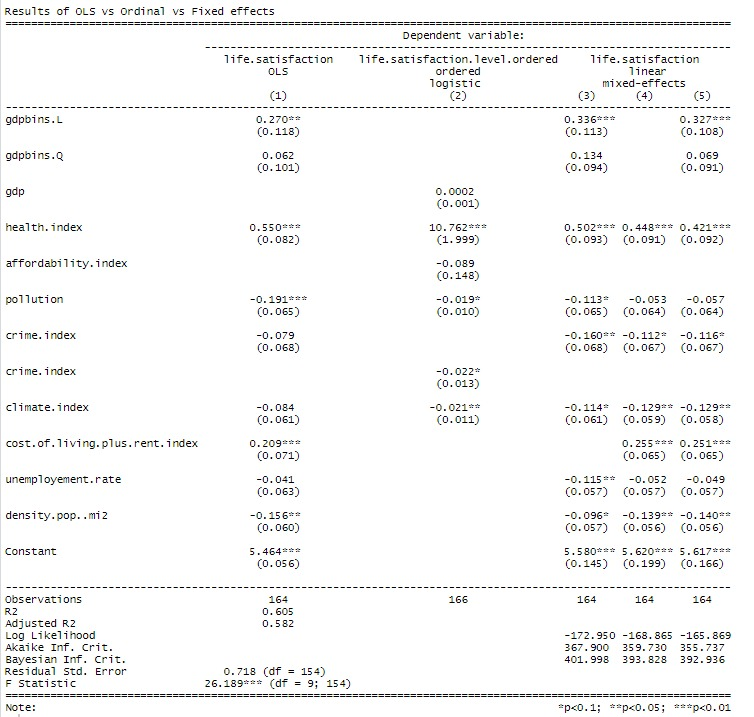
We included random effects using CLMM2 package.



We have converted GDP into bins and used it as a fixed effect and used continents as random effects.

1. GDP has a positive slope, which implies growth in GDP of each country leads to better life satisfaction rate.
2. With increase in health facilities in each country there is a chance of increase in the life satisfaction rate.
3. Pollution and life satisfaction rate are indirectly proportional.
4. With rise in crime rate, unemployment rate there is a fall in the life satisfaction rate.
5. In this model, we can see significant values for GDP, health index, crime index and unemployment rate

**Multilevel Models**



In these models, we have used numerical value of life satisfaction rate instead of categorical values

**We could see consistent significant values for few variables across all the models**.

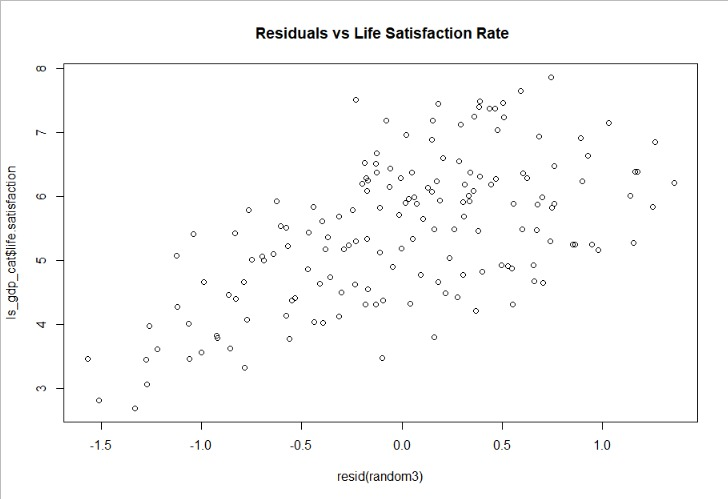
* We have converted GDP into bins and used it as a fixed effect and used continents as random effects.
* With increase in GDP, the life satisfaction rate is expected to increase.
* Health facilities has a direct effect on the life satisfaction rate.
* Pollution, Climate & Crime rate have a negative slope i.e. growth in either of them or all worsens the life satisfaction rate.
* Cost of living plus rent index is directly proportional to the life satisfaction rate of that country.
* With rise in unemployment rate & population density there is a fall in the life satisfaction rate.

1. **Quality Checks**

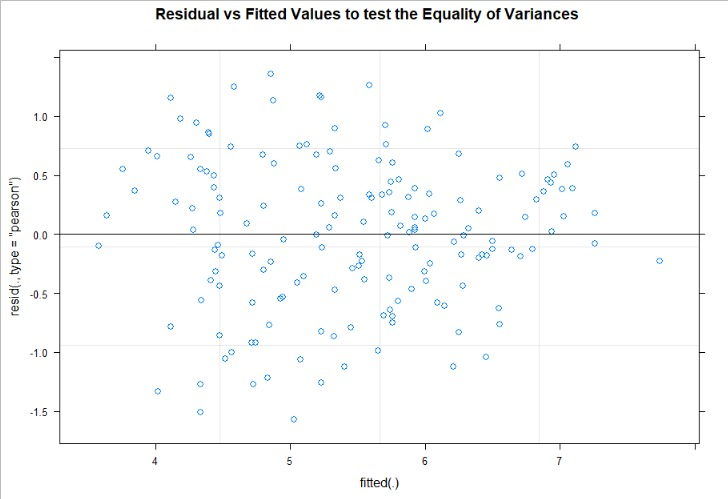
--Test of assumptions for our best model that has taken continent to be the random effect and GDP to be the fixed effect

In fixed effect model, we assume the error term to be normal.

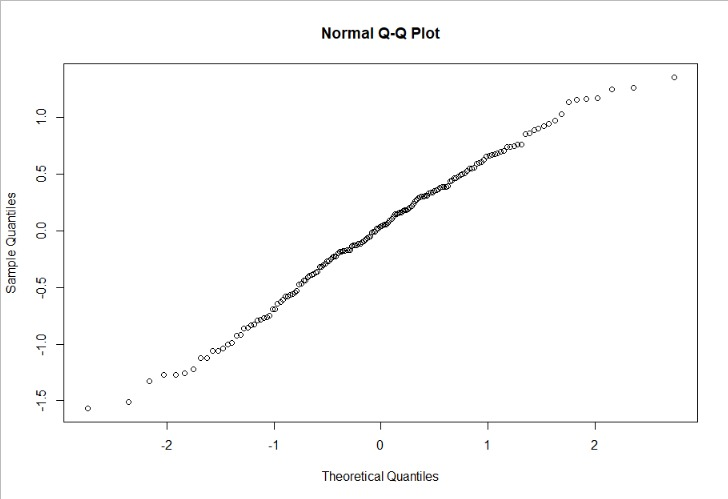
**Testing the assumptions of Linearity, Normality & Equality of Variances**



From the above plot, we can see that there is some sort of linear relation between the residuals and the life satisfaction rate.



Here we observe that the residuals are equally distributed around the centered line, but we can a few points above the centered line at the end of the distribution, that is because of few observations in that region.



The residual seems to be following the normal distribution for most part of the plot, but we see that there is a slight deviation at the tails.

Thus, we conclude saying that the residuals follow the LINE assumptions.

1. **Recommendations**

* Health index is the most significant and the strongest variable contributing to a better Life satisfaction rate. The government has to take appropriate measures and ensure that better health facilities are provided to its citizens.
* The next main factor to focus on is the GDP. Production of its own products and services within the country will not only increase the economy of the country but also provide better employment rates thereby providing the individuals with a better quality of life.
* Hotter climatic conditions have a negative impact on the life satisfaction rate. The government has to take precautionary measures to balance the ecosystem by growing more trees, reducing the emission of harmful gases into the air, making sure that factories and industries are established in areas which is away from cities or areas where human presence is less, making sure that the industry waste is not dumped into water bodies which can kill the aquatic balance and also eradicate the use of certain harmful waste like plastics.
* And lastly, focus on the safety of the individuals by imposing strict rules thereby reducing the crime rate.

1. **References**

<https://cran.r-project.org/web/packages/ClustImpute/index.html>

<https://stats.idre.ucla.edu/r/dae/ordinal-logistic-regression/>

<https://cran.r-project.org/web/packages/ordinal/vignettes/clmm2_tutorial.pdf>

<https://cran.r-project.org/web/packages/ordinal/ordinal.pdf>

Professor Daniel Zantedeschi, Assistant Professor in ISDS

Information Systems and Decision Sciences Department

Muma College of Business, University of South Florida, Tampa, FL , Email: [danielz@usf.edu](mailto:danielz@usf.edu)

1. **Appendix: R Code**

rm(list=ls())

getwd()

setwd("SDM")

library(rio)

library(stargazer)

library(moments)

library(car)

library(corrplot)

library(readxl)

**# Importing the data**

life = import("Project\_data.xlsx",sheet = "Lifesatisfaction")

hhincome = import("Project\_data.xlsx",sheet = "HouseholdIncome")

gdp = import("Project\_data.xlsx",sheet = "GDP")

health = import("Project\_data.xlsx",sheet = "Health Index")

crime = import("Project\_data.xlsx",sheet = "Crime Rate")

qol = import("Project\_data.xlsx",sheet = "Quality of Life")

unemp = import("Project\_data.xlsx",sheet = "Unemployment")

livingcost = import("Project\_data.xlsx",sheet = "Cost of Living")

pollution = import("Project\_data.xlsx",sheet = "Pollution")

pop = import("Project\_data.xlsx",sheet = "Population Density")

afford = import("Project\_data.xlsx",sheet = "Affordability")

continent = import("Project\_data.xlsx",sheet = "Continent")

**# Column names to lowercase**

colnames(life)=tolower(make.names(colnames(life)))

colnames(hhincome)=tolower(make.names(colnames(hhincome)))

colnames(gdp)=tolower(make.names(colnames(gdp)))

colnames(health)=tolower(make.names(colnames(health)))

colnames(crime)=tolower(make.names(colnames(crime)))

colnames(qol)=tolower(make.names(colnames(qol)))

colnames(unemp)=tolower(make.names(colnames(unemp)))

colnames(livingcost)=tolower(make.names(colnames(livingcost)))

colnames(pollution)=tolower(make.names(colnames(pollution)))

colnames(pop)=tolower(make.names(colnames(pop)))

colnames(afford)=tolower(make.names(colnames(afford)))

colnames(continent)=tolower(make.names(colnames(continent)))

**# Merging a dataset and reindexing the continent column to the left**

lifesatisfaction <- merge(life,continent, by = "country", all.life = TRUE)

#View(lifesatisfaction)

lifesatisfaction <- lifesatisfaction[, c(3,1,2)]

#View(lifesatisfaction)

summary(ls)

**# Merging all the data sets using left join. Join condition is by country and we used left join to get the values of all the countries for which**

**#we have the life satisfaction rate**

library(tidyverse)

ls = list(lifesatisfaction,hhincome,gdp,health,crime,qol,unemp,livingcost,pollution,pop,afford) %>% reduce(left\_join, by = "country")

#View(ls)

**#Structure of the dataframe**

str(ls)

**#Counting the NA values in each column**

map(ls, ~sum(is.na(.)))

**# Making the response variable categorical**

ls$life.satisfaction.level <- ifelse(ls$life.satisfaction<4, "Low","other")

ls$life.satisfaction.level <- ifelse(ls$life.satisfaction.level== "other" & ls$life.satisfaction >=4 & ls$life.satisfaction <6 , "Medium",ls$life.satisfaction.level)

ls$life.satisfaction.level <- ifelse(ls$life.satisfaction.level== "other" & ls$life.satisfaction >=6 , "High",ls$life.satisfaction.level)

#View(ls)

**#Countries in each category based on the Life Satisfaction Rate**

library(plyr)

lslevels <- table(ls$life.satisfaction.level)

plot(lslevels, main = " Countries in each categor based on the Life Satisfaction Rate ")

**# Reindexing the columns**

ls <- ls[, c(1,2,3,20,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19)]

#View(ls)

#ls$life.satisfaction.level = factor(ls$life.satisfaction.level)

ls$gdp <- ls$gdp/10^10

**# Data Visualizations and Descriptive Analysis**

ls[which.min(ls$health.index),]

par(mfrow=c(2,3))

plot(ls\_gdp\_cat$life.satisfaction~ ls\_gdp\_cat$crime.index.,main = " Life Satisfaction Rate vs Crime Index")

plot(ls\_gdp\_cat$life.satisfaction~ ls\_gdp\_cat$health.index,main = " Life Satisfaction Rate vs Health Index")

plot(ls\_gdp\_cat$life.satisfaction~ ls\_gdp\_cat$pollution,main = " Life Satisfaction Rate vs Pollution")

plot(ls\_gdp\_cat$life.satisfaction~ ls\_gdp\_cat$unemployement.rate,main = " Life Satisfaction Rate vs Unemployment Rate")

plot(ls\_gdp\_cat$life.satisfaction~ ls\_gdp\_cat$climate.index,main = " Life Satisfaction Rate vs Climate Index")

plot(ls\_gdp\_cat$life.satisfaction~ ls\_gdp\_cat$affordability.index,main = " Life Satisfaction Rate vs Population Density")

**# Checking the null values in each variable**

#install.packages("mice")

#install.packages("VIM")

library(VIM)

library(mice)

md.pattern(ls)

aggr\_plot <- aggr(ls, col=c('navyblue','red'), numbers=TRUE, sortVars=TRUE, labels=names(ls),ylim = c(0,0.5), cex.axis=.7, gap=3, ylab=c("Histogram of missing data","Pattern"))

**#K MEANS CLUST IMPUTE for missing values**

**#Extracting the numerical columns for Clust Impute**

ls1 = subset(ls, select = -c(1,2,3,4) )

lifesatisfaction\_cat = subset(ls, select = c(1,2,3,4) )

#View(ls1)

#View(lifesatisfaction\_cat)

nr\_iter <- 10 # iterations of procedure

n\_end <- 10 # step until convergence of weight function to 1

nr\_cluster <- 7# number of clusters

c\_steps <- 50 # numer of cluster steps per iteration

#install.packages("ClustImpute")

library(ClustImpute)

ls2 <- ClustImpute(ls1,nr\_cluster=nr\_cluster, nr\_iter=nr\_iter, c\_steps=c\_steps, n\_end=n\_end, wf = default\_wf)

View(ls2)

str(ls2)

class(ls2)

#View(ls2$complete\_data)

ls\_full <- cbind(lifesatisfaction\_cat,ls2$complete\_data)

#View(ls\_full)

aggr\_plot <- aggr(ls\_full, col=c('navyblue','red'), numbers=TRUE, sortVars=TRUE, labels=names(ls\_full), cex.axis=.7, gap=3, ylab=c("Histogram of missing data","Pattern"))

**#Correlation plot**

xx=cor(ls2$complete\_data)

corrplot(xx,method="ellipse",type="lower")

corrplot(xx,method="number",type="lower")

**#Ordinal Logistic Regression**

View(ls\_full)

ls\_full$life.satisfaction.level.ordered <- factor(ls\_full$life.satisfaction.level, levels=c("Low", "Medium", "High"), ordered=TRUE)

library(MASS)

attach(ls\_full)

ordinal1 <- polr(life.satisfaction.level.ordered ~ gdp + climate.index + pollution +ls\_full$crime.index+health.index+

ls\_full$affordability.index , data=ls\_full, Hess=TRUE)

summary(ordinal1)

stargazer(ordinal1,type="text")

ls\_gdp\_cat <- ls\_full

#install.packages("binr")

library(binr)

#install.packages("dlookr")

library(dlookr)

**#Quantile based binning is a good strategy to use for adaptive binning.**

**#Quantiles are specific values or cut-points which help in partitioning the continuous valued distribution of a**

**#specific numeric field into discrete contiguous bins or intervals.**

ls\_gdp\_cat$gdpbins <- binning(ls\_gdp\_cat$gdp, nbins = 3,

labels = c("Low", "Medium", "High"), type = "quantile")

View(ls\_gdp\_cat)

ls\_gdp\_cat <- ls\_gdp\_cat %>% filter(country != c('United States')) %>% filter(country

!= c('China'))

**#Standardization of values for better training**

ls\_gdp\_cat <- ls\_gdp\_cat %>% mutate\_at(c("medianpercapitaincome","medianhouseholdincome","health.index", "gdp","crime.index.","safety.index","traffic.commute.time.index",

"climate.index",

"unemployement.rate","cost.of.living.plus.rent.index","groceries.index","restaurant.price.index",

"pollution","density.pop..mi2","affordability.index","population"), ~(scale(.) %>% as.vector))

**#Ordinal logistic Regression with Random effects**

install.packages("ordinal")

library(ordinal)

attach(ls\_gdp\_cat)

fm1 <- clmm2(life.satisfaction.level.ordered ~ continent +health.index + pollution + crime.index.+

unemployement.rate+ affordability.index+density.pop..mi2, random = gdpbins, data=ls\_gdp\_cat, Hess=TRUE)

summary(fm1)

ls\_gdp\_cat$continent <- as.factor(ls\_gdp\_cat$continent)

fm2 <- clmm2(life.satisfaction.level.ordered ~ gdpbins +health.index + pollution + crime.index.+

unemployement.rate+ affordability.index+density.pop..mi2, random = continent, data=ls\_gdp\_cat, Hess=TRUE)

summary(fm2)

levels(ls\_gdp\_cat$continent)

View(ls\_gdp\_cat)

**# Pool MOdel vs Multilevel models**

library(lme4)

lm <- lm(life.satisfaction ~ gdpbins + health.index + pollution +ls\_gdp\_cat$crime.index+climate.index+cost.of.living.plus.rent.index+

unemployement.rate+ density.pop..mi2 , data=ls\_gdp\_cat)

summary(lm)

random1 <- lmer(life.satisfaction ~ gdpbins+ health.index + pollution +ls\_gdp\_cat$crime.index+climate.index+

unemployement.rate + density.pop..mi2+ (1+1|continent), data=ls\_gdp\_cat, REML=FALSE)

summary(random1)

random2 <- lmer(life.satisfaction ~ health.index + pollution +ls\_gdp\_cat$crime.index+climate.index+cost.of.living.plus.rent.index+

unemployement.rate+ density.pop..mi2+ (1|continent)+ (1+1|gdpbins), data=ls\_gdp\_cat, REML=FALSE)

summary(random2)

random3 <- lmer(life.satisfaction ~ gdpbins + health.index + pollution +ls\_gdp\_cat$crime.index+climate.index+

cost.of.living.plus.rent.index+

unemployement.rate+ density.pop..mi2+ (1|continent) , data=ls\_gdp\_cat, REML=FALSE)

summary(random3)

stargazer(lm,ordinal1,random1,random2,random3, title="Results of OLS vs Ordinal vs Fixed effects", align=TRUE,type="text")

ls\_da1 = subset(ls\_gdp\_cat, select = -c(1,2,6,5,4,7,11,12,16,17) )

summary(ls\_da)

ls\_da1 <- subset(ls\_da1,select = -c(11,12))

labels <- paste(c("Life satisfaction",

"GDP",

"Health Index",

"Crime Index",

"Climate Index",

"Unemployment Rate",

"Cost of Living",

"Pollution",

"Population Density",

"Affordability Index" ))

boxplot(ls\_da1, xaxt = "n", xlab = "")

**# x axis with ticks but without labels**

axis(1, labels = FALSE)

**# Plot x labs at default x position**

text(x = seq\_along(labels), y = par("usr")[3] - 1, srt = 30, adj = 1,

labels = labels, xpd = TRUE)

**# Quality Checks**

plot(random3, main = " Residual vs Fitted Values to test the Equality of Variances")

ggplot(data.frame(eta=predict(random3,type="link"),pearson=residuals(random3,type="pearson")),

aes(x=eta,y=pearson)) +

geom\_point() +

theme\_bw()

qqnorm(residuals(random3))

abline(0,1)

plot(resid(random3),ls\_gdp\_cat$life.satisfaction,main = " Residuals vs Life Satisfaction Rate")

plot(factor(ls$country),ls$crime.index.)