Ines Rondeau Petit
Data Analytics BootCamp
September 2021

World Happiness Report

Mid-course data analytics project

Official variables

X *(Variables)*

Objective

GDP per capita WDI (World Development Indictor)
Healthy Life Expectancy WHO (World Health Organization)

Subjective

(Average values for answers 0/1)

Social Support *("If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?")*
Freedom to make life choices *("Are you satisfied or dissatisfied with your freedom to choose what you do with your life?)*
Generosity *("Have you donated money to a charity in the past month?")*
Corruption Perception *("Is corruption widespread throughout the government or not" and "Is corruption widespread within businesses or not?")*

Official variables

y *(targuet)*

Subjective

(Average values for answers between 0-10)

Happiness score *('Life Ladder')*

ID

String

(object dtype not used for model)

Country name

Source: [worldhappiness.report] *'The World Happiness Report 2015'*

Official model

LinearRegression()

Targuet Value in a float data type *(‘Life Ladder’)*

r2_score()

0.738

mean_squared_error()

0.366

Project goals

(1) Create an accurate model changing the subjective variables for objective ones

(2) Test two hypothesis:

Is happines is related to religion? and what about alcohol consumption?

New variables

Education

Mostly correlated to **Social Support** and **Freedom to make life choices**

/float dtypes

lu (% of No Schooling Attained in Population)
lpc (% of Complete Primary Schooling Attained in Popuplation)
lsc (% of Complete Secondary Schooling Attained in Population)
lhc (% of Complete Tertiary Schooling Attained in Population)
yr_sch (Average Years of Schooling Attained)

Source: [barrolee.com] *'Barro_Lee Educational Attainment Dataset'*

New variables

Gender equality

Mostly correlated to **Social Support** and **Freedom to make life choices**

/float dtypes

Adolescent fertility rate (births per 1,000 women ages 15-19)
Employers, female (% of female employment)
Labor force, female (% of total labor force)

/int dtypes (0=no/1=yes)

The government administers maternity leave benefits
Paid parental leave
Legislation addressing domestic violence
The law prohibits discrimination to access credit based on gender
Law prohibits discrimination on employiment based on gender

Source: [databank.worldbank.org] *'Gender statistics'*

New variables

Corruption

Mostly correlated to **Perception of corruption** and **Freedom to make life choices**

/int dtypes

CPI 2015 Score (Perceived levels of corruption according to experts, uses a scale of 0 to 100, where zero is highly corrupt and 100 is very clean)

Source: [transparency.org] *'Corruption Perception Index'*

New variables

‘Giving’ scores

Mostly correlated to **Generosity** and **Freedom to make life choices**

/int dtypes

Score(%)_hs (Rounded percentge of adults that have often helped a stranger in a needed situation)
Score(%)_d (Rounded percentage of people that has donated money to a charity)
Score(%)_v (Rounded percentage of people that has voluteered their time)

Source: [cafonline.org] *‘World Giving Data’*

Final new variables

The ones I kept that actually improved my model

- CPI 2015 Score
- Isc
- Adolescent fertility rate
- Labor force, female
- The government administers maternity leave benefits
- Legislation addressing domestic violence
- Score(%)_hs
- Score(%)_v

- Log GDP per capita
- Healthy life expectancy at birth

New model

LinearRegression()

Targuet Value in a float data type *('Life Ladder')*

r2_score()

0.716

mean_squared_error()

0.316

Hypothesis testing

Bring data

Religion /float dtype
yes_percentage (% of people that aswered yes to *'is religion important?'*)

Alcohol abuse /float dtype
value (Alcohol consumption per capita (15 years+) in litres of pure alcohol

Tresholders for comparison

Mean happiness scores (5.4)

Devided the DataFrames in two: one with 'happy' scores (>5.5), and 'sad' scores (<= 5.5)

Source: [who.int] *'Levels of Consumption'*
[gallup.com] *'Importance of religion by country'*

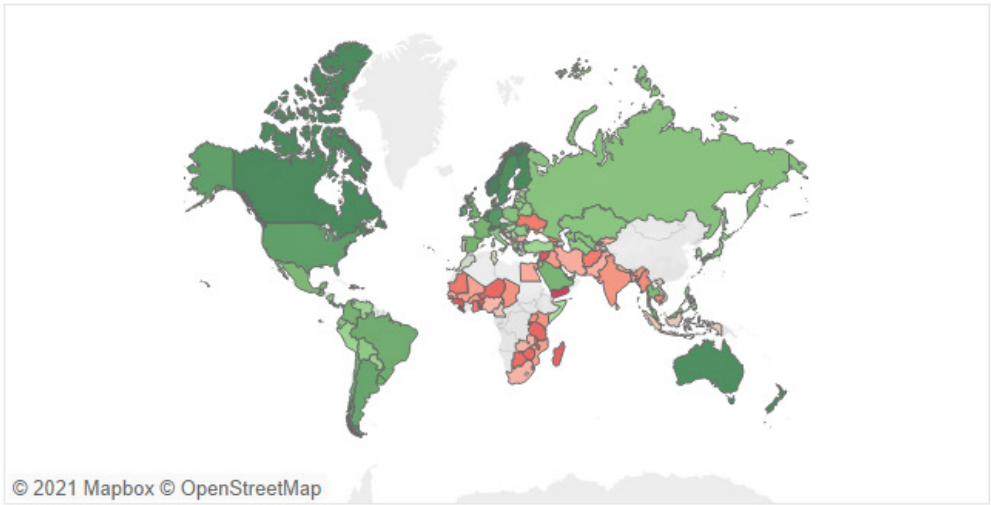
Hypothesis 1

Are religion and happiness dependent variables?

H0: religion importance is equal for countries with high and low happiness scores

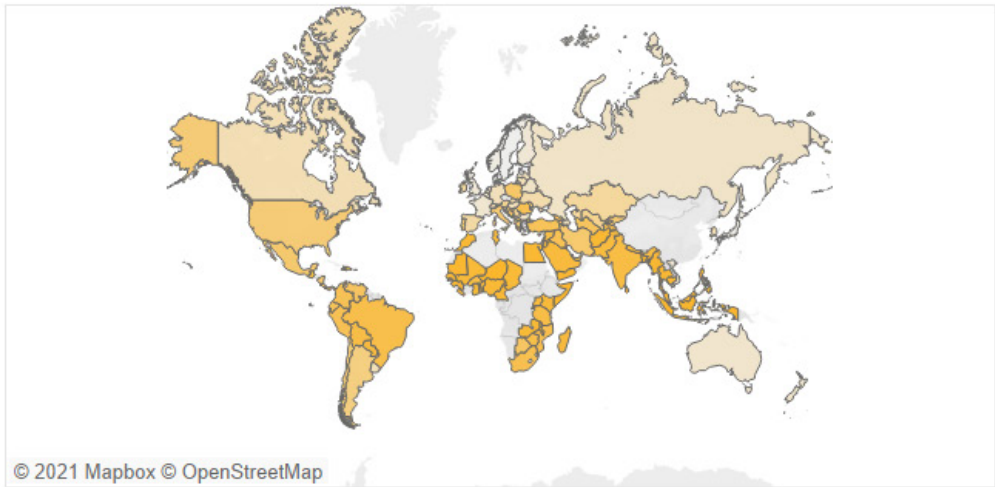
p_value = 2.0804627004984745e-27

H0 rejected, they're **DEPENDENT** values



happiness score

2,702 — 7,603

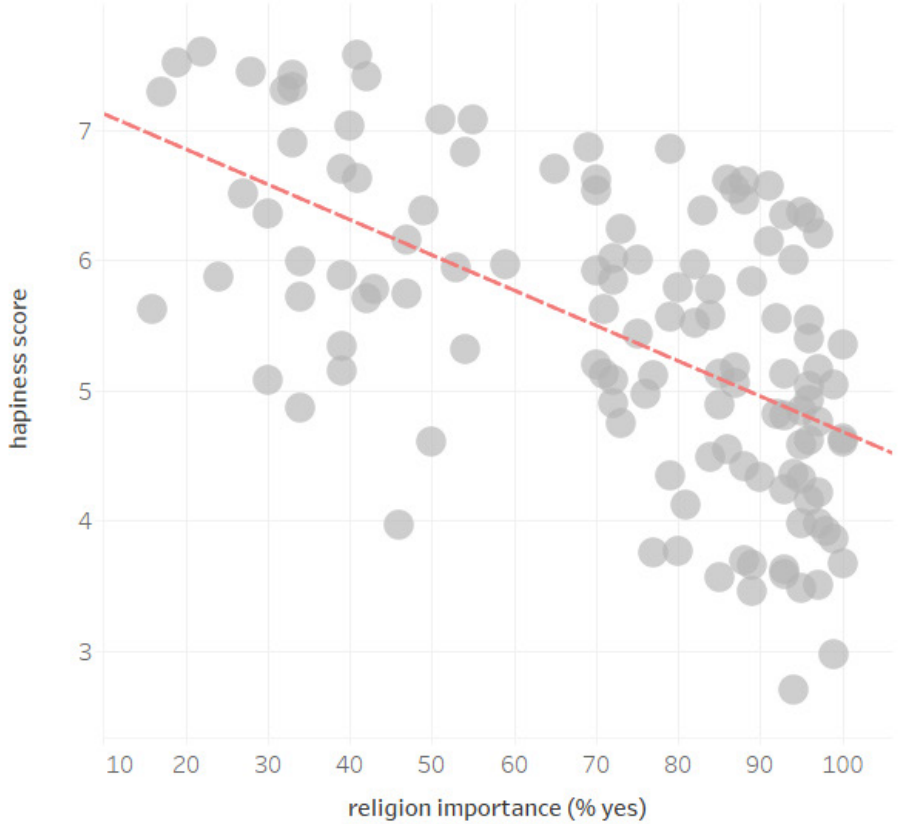


religion importance

16,00 — 100,00

ScatterPlot

Religion importance x Hapiness Score



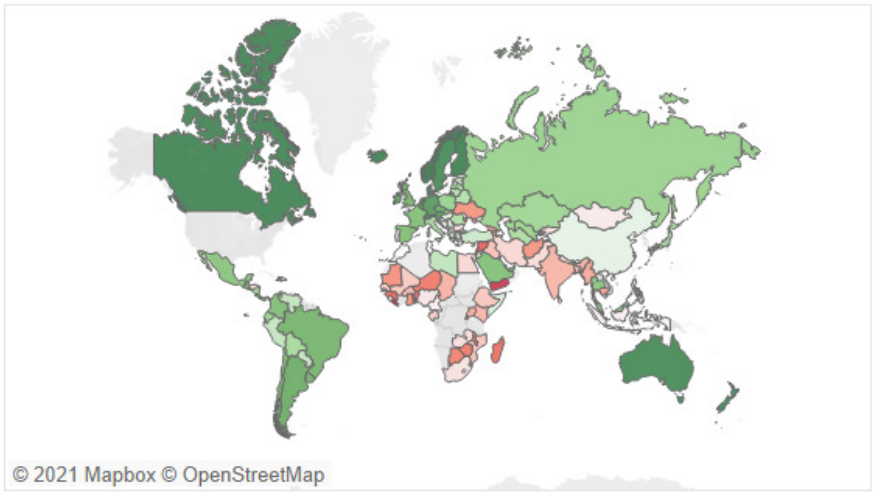
Hypothesis 2

Are alcohol consumption and happiness dependent variables?

H0: alcohol abuse is equal for countries with high and low happiness scores

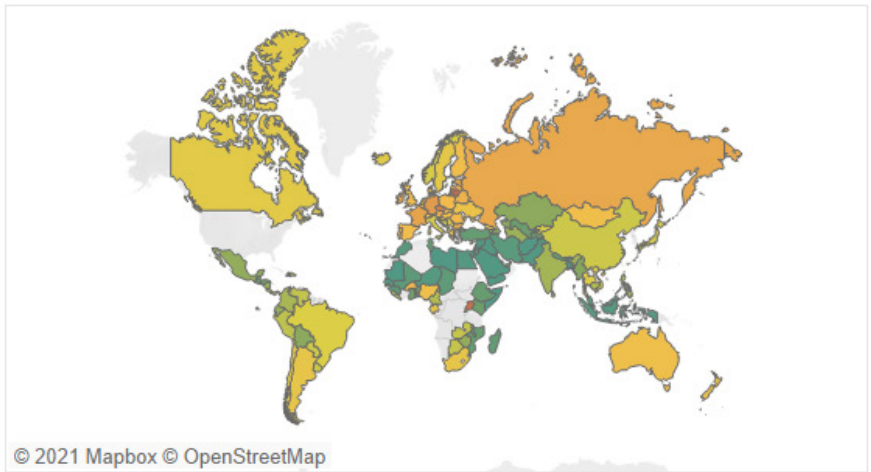
p_value = 0.0027612039650190302

H0 rejected, they're **DEPENDENT** values



Happiness score

2,702 — 7,603

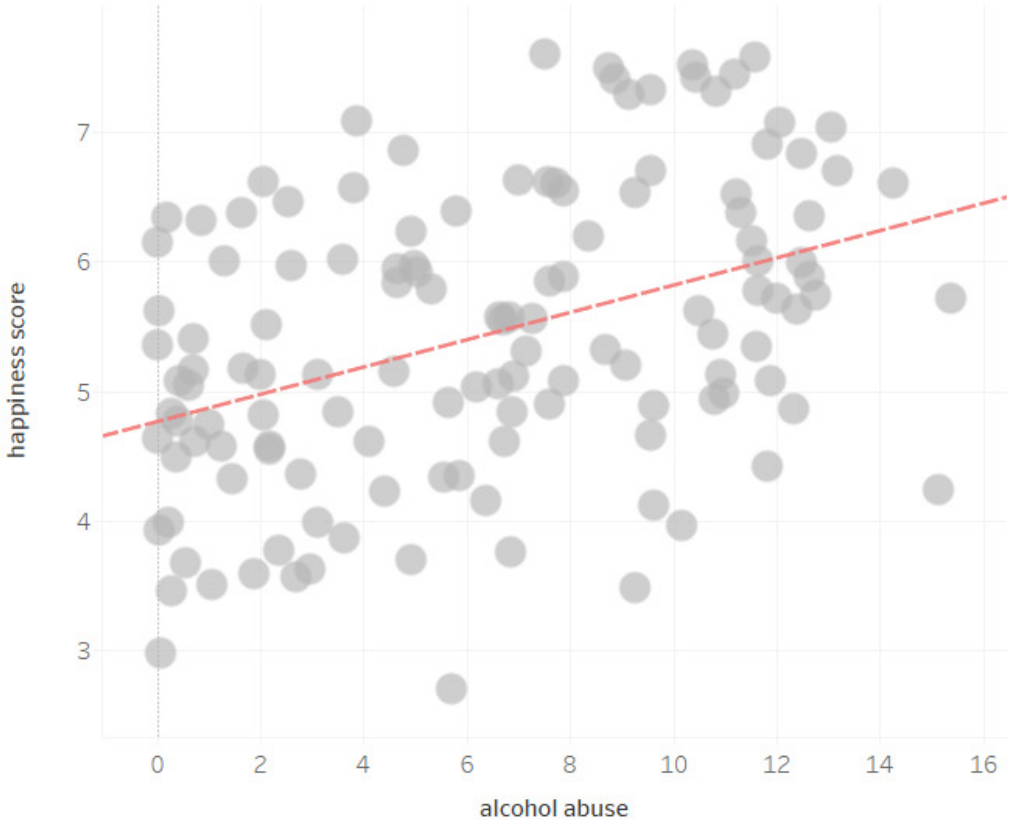


Alcohol abuse

0,00 — 15,38

ScatterPlot

Alcohol Abuse x Hapiness Score



Conclusions

/It's feasable to create an accurate model based solely on objective data.

/The 'happiest' the country, the more alcohol they consume, generally.

/Religion importnce implies, in most cases, lower happiness scores.