

Analiza faktora sreće po državama

Daniel Katić, Lana Kohut

2025-11-12

Uvod

Ovaj projekt analizira povezanost ekonomskih i društvenih faktora sa srećom stanovništva koristeći podatke iz World Happiness Reporta (2016). Izvešće obuhvaća niz pokazatelja koji zajedno tvore tzv. indeks sreće — numerički pokazatelj koji procjenjuje koliko su stanovnici pojedine države zadovoljni svojim životom.

Cilj ove analize je istražiti koliko ekonomski pokazatelji, poput BDP-a po stanovniku, te društveni čimbenici poput obiteljske podrške, slobode izbora i povjerenja u institucije, doprinose ukupnoj percepciji sreće u različitim državama svijeta. Korištenjem osnovnih statističkih mjera (aritmetička sredina, medijan, mod, raspon, varijanca, standardna devijacija, percentili i kvartili), te analizom korelacije između ključnih varijabli, moguće je steći bolji uvid u odnose između ekonomskog blagostanja i subjektivnog osjećaja sreće.

Osim deskriptivne statistike, projekt vizualno prikazuje distribuciju vrijednosti i međusobne povezanosti podataka radi lakšeg razumijevanja rezultata. Dobiveni uvidi mogu poslužiti kao temelj za dublje analize u područjima ekonomske politike, sociologije i psihologije blagostanja.

Projekt je izrađen u okruženju RStudio, koristeći R Markdown i programski jezik R za obradu, analizu i vizualizaciju podataka. Autori projekta su Daniel Katić (0303123347) i Lana Kohut.

Uvoz podataka

```
# Uvoz potrebnih biblioteka
suppressPackageStartupMessages({
  library(readr)
  library(dplyr)
})

# Podatke učitavamo u varijablu data_2016
raw_data_2016 <- read_csv("podaci/2016.csv")
spec(raw_data_2016)

## cols(
##   Country = col_character(),
##   Region = col_character(),
##   `Happiness Rank` = col_double(),
##   `Happiness Score` = col_double(),
##   `Lower Confidence Interval` = col_double(),
##   `Upper Confidence Interval` = col_double(),
##   `Economy (GDP per Capita)` = col_double(),
##   Family = col_double(),
##   `Health (Life Expectancy)` = col_double(),
##   Freedom = col_double(),
##   `Trust (Government Corruption)` = col_double(),
##   Generosity = col_double(),
##   `Dystopia Residual` = col_double()
## )

# Dodajemo stupac Godina
data_2016 <- raw_data_2016 %>% mutate(Godina = 2016)
```

Opis stupaca u datasetu *World Happiness Report 2016*

Table 1: Opis stupaca u datasetu

Naziv.stupca	Opis	Tip.podatka	Primjer
Country	Naziv države	Karakter (tekst)	Denmark
Region	Geografska regija u kojoj se država nalazi	Karakter (tekst)	Western Europe
Happiness Rank	Rang države prema indeksu sreće (1 = najsretnija država)	Numerički (integer)	1
Happiness Score	Ukupni indeks sreće za državu (0–10 skala, veća vrijednost = sretnija država)	Numerički (decimalni)	7.526
Lower Confidence Interval	Donja granica intervala pouzdanosti za Happiness Score	Numerički (decimalni)	7.460
Upper Confidence Interval	Gornja granica intervala pouzdanosti za Happiness Score	Numerički (decimalni)	7.592
Economy (GDP per Capita)	Procjena utjecaja BDP-a po stanovniku na Happiness Score (0–1 skala)	Numerički (decimalni)	1.44178
Family	Procjena utjecaja obiteljske i društvene podrške (0–1 skala)	Numerički (decimalni)	1.16374
Health (Life Expectancy)	Procjena utjecaja očekivanog životnog vijeka na sreću (0–1 skala)	Numerički (decimalni)	0.79504
Freedom	Procjena koliko su građani slobodni u donošenju životnih odluka (0–1 skala)	Numerički (decimalni)	0.57941
Trust (Government Corruption)	Povjerenje građana u vladu / razina percipirane korupcije (0–1 skala)	Numerički (decimalni)	0.44453
Generosity	Procjena razine velikodušnosti u društvu (npr. donacije, pomaganje drugima)	Numerički (decimalni)	0.36171
Dystopia Residual	Dodatna komponenta modela koja predstavlja ostatak bodova sreće (referentna vrijednost “najgorih uvjeta”)	Numerički (decimalni)	2.73939

Opisna statistika

Prikazati ćemo aritmetičku sredinu, medijan, i ostale na “Happiness Score” stupcu.

Aritmetička sredina

Aritmetička sredina je srednja vrijednost podataka, odnosno zbroj svih elemenata podijelimo sa brojem elemenata.

$$\bar{x}_n = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

```
arit_sred <- mean(data_2016$'Happiness Score', na.rm = TRUE)
```

$\bar{x} = 5.3821847$

Medijan

Medijan je središnja vrijednost skupa podataka. Da bismo ga izračunali, potrebno je najprije sortirati podatke od najmanje do najveće vrijednosti.

Ako skup podataka ima neparan broj elemenata, medijan je jednostavno srednji element.

Ako skup ima paran broj elemenata, medijan se računa kao prosjek dviju srednjih vrijednosti, prema formuli:

$$\frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

```
medijan <- median(data_2016$'Happiness Score', na.rm = TRUE)
```

Medijan = 5.314

Mod

Mod je podatak koji se najčešće pojavljuje u podatkovnom skupu. On ne mora nužno postojati, jer možemo imati skup brojeva i svaki broj se pojavljuje samo jednom.

```
get_mode <- function(v) {  
  uniqv <- unique(v)  
  uniqv[which.max(tabulate(match(v, uniqv)))]  
}  
mod <- get_mode(data_2016$'Happiness Score')
```

Mod = 6.379

Mjere raspršenja podataka