

# ANA 515 Assignment 2

Saif Hossain

2022-09-16

## Fifa

The data here are countries that play soccer and their attributes. The countries that play soccer, in which confederation they play for example Europe plays in Uefa. Population\_share, tv\_audience\_share and gdp\_weighted share. The data measures TV audience share for marketing purposes. The data is opensource data and can be collected from public source. The data can be used to allocate money for advertisement. The country with bigger tv audience share greater the cost of advertising in that country. The data is stored in csv (comma separated values) format. The data frame is delimited, comma is the delimiter. If it was a binary file, binary editor could be used to open it or powershell could be used as well. Since, this is a csv file Excel will most commonly be used to open the file.

```
#Reading the data into R.

# echo = FALSE prevents code, but not the results from appearing in the finished file. Use this when writing to a file.

url <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/fifa/fifa_countries_audience.csv"

fifa_data <- read.csv(url) #Assigning the url to object

# Cleaning Data using functions.
fifa_data <- rename(fifa_data, Audience_Percent = 'tv_audience_share')
fifa_data <- rename(fifa_data, Country = 'country')
fifa_data <- rename(fifa_data, Confederation = 'confederation')
fifa_data <- rename(fifa_data, Population_Percent = 'population_share')
fifa_data <- rename(fifa_data, GDP_Weighted_Share = 'gdp_weighted_share')
fifa_data <- filter(fifa_data, Audience_Percent > 0.0) #Filtering out data where there is 0 TV audience
head(fifa_data)

# Getting columns of my dataframe

colnames(fifa_data)

## [1] "Country"          "Confederation"      "Population_Percent"
## [4] "Audience_Percent" "GDP_Weighted_Share"

# Characteristics of the data. And Inline code below.

dimensions <- dim(fifa_data)
```

Table 1: Fifa Advertising Data

Columns	Summary
Country	Country Playing Soccer.
Audience_Percent	Percentage of people watching Soccer.
Confederation	The league they are playing.
Population_Percent	What is the percentage of people vs who is watching soccer.
GDP_Weighted_Share	GDP of the country in terms of watching soccer.

This dataframe has 124 rows and 5 columns. The names of the columns and a brief description of each are in the table below:

```
# Column Description using kable

col_des <- data.frame(
  Columns = c("Country", "Audience_Percent", "Confederation", "Population_Percent", "GDP_Weighted_Share"),
  Summary = c("Country Playing Soccer.", "Percentage of people watching Soccer.", "The league they are
            ")
)

kbl(col_des, caption = "Fifa Advertising Data") %>%
  kable_styling(bootstrap_options = c("striped", "hover"))
```

#Summary statistics for columns without missing values.

```
fifa_data_aud <- filter(fifa_data, Audience_Percent > 0.3) #Filtering out data where there is greater than 30% audience
fifa_data_pop <- filter(fifa_data, Population_Percent > 0.2) #Filtering out data where there is greater than 20% population
fifa_data_gdp <- filter(fifa_data, GDP_Weighted_Share > 0.0) #Filtering out data where there is greater than 0.0 GDP share

a <- summary(fifa_data_aud$Audience_Percent)

min_aud <- min(fifa_data_aud$Audience_Percent)
max_aud <- max(fifa_data_aud$Audience_Percent)
mean_aud <- mean(fifa_data_aud$Audience_Percent)

a1 <- sum(is.na(fifa_data_aud$Audience_Percent)) #Number of missing values in Audience_percent column

mis_aud <- a1

b <- summary(fifa_data_aud$Population_Percent)

min_pop <- min(fifa_data_aud$Population_Percent)
max_pop <- max(fifa_data_aud$Population_Percent)
mean_pop <- mean(fifa_data_aud$Population_Percent)

b1 <- sum(is.na(fifa_data_aud$Population_Percent)) #Number of missing values in Population_Percent column
```

Table 2: Fifa Profitable Country Data

Summary	Minimum	Maximum	Mean	Missing
Audience Percentage	0.4	14.8	1.977273	0
Population Percentage	0.4	14.8	1.977273	0
GDP Weighted Share	0.0	11.3	2.002273	0

```

mis_pop <- b1

c <- summary(fifa_data_aud$GDP_Weighted_Share)

min_gdp <- min(fifa_data_aud$GDP_Weighted_Share)
max_gdp <- max(fifa_data_aud$GDP_Weighted_Share)
mean_gdp <- mean(fifa_data_aud$GDP_Weighted_Share)

c1 <- sum(is.na(fifa_data_aud$GDP_Weighted_Share)) #Number of missing values in GDP_Weighted_Share column

mis_gdp <- c1

summary_fifa <- data.frame(
  Summary = c("Audience Percentage", "Population Percentage", "GDP Weighted Share"),
  Minimum = c(min_aud, min_pop, min_gdp),
  Maximum = c(max_aud, max_pop, max_gdp),
  Mean = c(mean_aud, mean_pop, mean_gdp),
  Missing = c(mis_aud, mis_pop, mis_gdp)

)

kbl(summary_fifa, caption = "Fifa Profitable Country Data") %>%
  kable_styling(bootstrap_options = c("striped", "hover"))

```