Assignment 2: Survey developpment

105.708 Data Acquisition and Survey Methods

Group 3: Gjorgieva Aleksandra(), de Lambertye Grégoire(12202211), Loos Annika Katharina()

2023-06-13

Contents

Introducction

topic

blabla

Hypothesis

blabla

question

blaba

Data exploration

```
file_path <- "./data/group_3.csv"

rawData <- read.csv(file_path, sep=";")
# rename the column in a english friendly format
colnames(rawData)[4] <- "answer_1"
colnames(rawData)[5] <- "answer_2"
colnames(rawData)[6] <- "answer_3"

print(head(rawData))
```

```
##   Gender Age             Academic.Program answer_1
## 1 female  23 Data Science / MSc / TU Wien      120
## 2   male  21 Data Science / MSc / TU Wien      180
## 3 female  23 Data Science / MSc / TU Wien       90
## 4 female  24 Data Science / MSc / TU Wien      140
## 5   male  22 Data Science / MSc / TU Wien      100
## 6   male  25                 Data Science       50
##                                  answer_2
## 1                                       4
## 2 5 (I think I should send way less time)
## 3 5 (I think I should send way less time)
## 4                                       4
## 5                                       4
## 6                                       4
##                                   answer_3
## 1                            YouTube video
## 2                            YouTube video
## 3 YouTube shorts - Tiktoks - Intsagram reels
## 4                         TV shows (series)
## 5                            YouTube video
## 6                            YouTube video
```

```
print(paste("The data set dimensions are ", dim(rawData)[1], " x ", dim(rawData)[2]))
```

```
## [1] "The data set dimensions are  38  x  6"
```

Preprocessing

The columnns Gender and Age look nice and don't requier preprocessing.

Academic programm

```
print(unique(rawData[3]))
```

```
##                                Academic.Program
## 1                    Data Science / MSc / TU Wien
```

```
## 6                                     Data Science
## 9               Business Informatics / BSc/ TU Wien
## 10                                MSc Data Science
## 11                     Data Science / MSc / TU WIen
## 12                                Data Science Msc
## 17 Statistik und Wirtschaftsmathematik BSc TU Wien
## 23                                Data Science MSc.
## 25                                              MSc
## 28                        Erasmus student, Statistic
## 34       Data Science / MSc / University of Zagreb
## 36             Business Informatics / BSc / TU Wien
```

As we see the format differs for the academic program answer. Where we have missing information we associate them by default to Data Science / MSc / TU Wien. To say we change "Data Science", "MSc Data Science", "MSc", "Data Science Msc", " Data Science MSc." to "Data Science / MSc / TU Wien".

```
row_index <- which(rawData$Academic.Program %in% c("Data Science",
                                                   "Data Science / MSc / TU Wien",
                                                   "MSc Data Science",
                                                   "Data Science / MSc / TU WIen",
                                                   "MSc",
                                                   "Data Science Msc",
                                                   "Data Science MSc."))

rawData[row_index, "Academic.Program"] <- "Data Science / MSc / TU Wien"
```

```
print(unique(rawData[3]))
```

```
##                                   Academic.Program
## 1                     Data Science / MSc / TU Wien
## 9               Business Informatics / BSc/ TU Wien
## 17 Statistik und Wirtschaftsmathematik BSc TU Wien
## 28                        Erasmus student, Statistic
## 34       Data Science / MSc / University of Zagreb
## 36             Business Informatics / BSc / TU Wien
```

```
rawData$answer_2 <- str_sub(rawData$answer_2, end = 2)
rawData$answer_2 <- as.numeric(rawData$answer_2)
```

```
for (i in 1:length(rawData$answer_1)){
  if(rawData$answer_1[i] <= 5){
    rawData$answer_1[i] <- rawData$answer_1[i] * 60
  }
}
```

```
head(rawData)
```

```
##   Gender Age             Academic.Program answer_1 answer_2
## 1 female  23 Data Science / MSc / TU Wien      120        4
## 2   male  21 Data Science / MSc / TU Wien      180        5
## 3 female  23 Data Science / MSc / TU Wien       90        5
```

```
## 4 female  24 Data Science / MSc / TU Wien      140        4
## 5   male  22 Data Science / MSc / TU Wien      100        4
## 6   male  25 Data Science / MSc / TU Wien       50        4
##                                     answer_3
## 1                                YouTube video
## 2                                YouTube video
## 3 YouTube shorts - Tiktoks - Intsagram reels
## 4                            TV shows (series)
## 5                                YouTube video
## 6                                YouTube video
```

Analysis

hypothesis 1

hypothesis 2

hypothesis 3

Conclusion