Final project

L

1. Introduction

In this study, a cross-sectional survey is conducted with a sample size of 1182 students of different age groups from different educational institutions in Delhi National Capital Region during covid 19 lockdown (NCR). Our general research question, is whether there is a correlation between the lockdown habits and mental health consequences. The problem we would like to solve, is to check whether the lockdown habits are related to mental and physical health, and which habits should we generally embrace in case of another lockdown, based on statistics and insights retrieved from the data. (Backed by related mental and physical health researches, pinned in Apendix.) We find these questions pretty valuable, since we are all related to lockdown experience, and defining healthy habits, can reduce a country resources invested on health services, as well as mental health services. Previous research and work can tell us which habits and symptoms are related to poor mental health. We'd like to use these hypotheses accordingly, and develop hypotheses of our own, which would to general recommendations of habits, to undergo lockdown in an healthy method both physically and mentally. Such questions can be difficult since the studies about lockdowns effects are still on-the-go since severe Covid distribution is still common around the globe, and such cases and general lockdown are still quite probable. Knowing it, we would like get a better evaluation of ideal approach for the general person, in terms of health and mental health measures. Our approach involves analyzing the general habits in lockdown, presenting its medical and mental consequences, based on research. After defining the mean and outliers, we would like to present the ideal habits. Our approach differs from other approaches by using Covid lockdowns as an extreme case, to generate answers on how much, and which 'not ideal' habits affect our general health and mental health.

2. Data

In this study, a cross-sectional survey is conducted with a sample size of 1182 students of different age groups from different educational institutions in Delhi National Capital Region (NCR). The survey was conducted as a one time survey around the mid-time of the public lockdown around this region, where each subject applied his subjective answers for the asked questions given. This survey presents lots of different features and habits during lockdown, and thus, would help us getting better understanding which feature influences general physical and mental health, and how. There are 19 features in this dataset, containing different questions regarding habits, and behaviors in lock down, where the features that seem most interesting and would be used during analysis are the following:
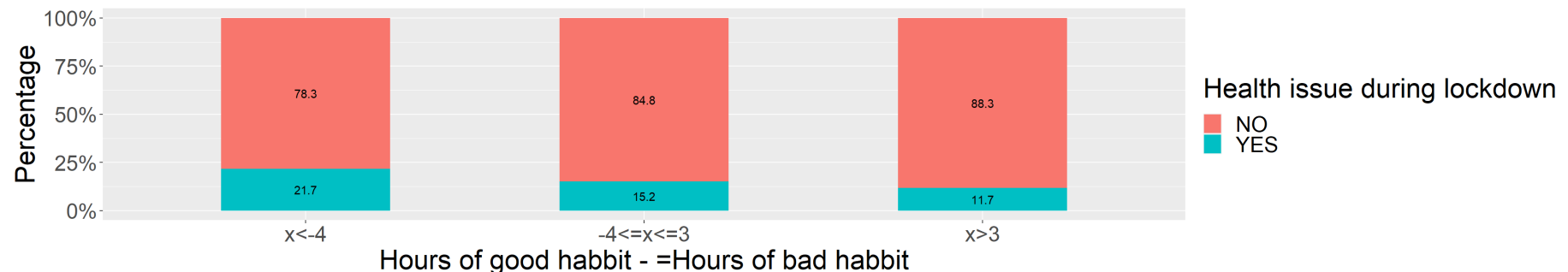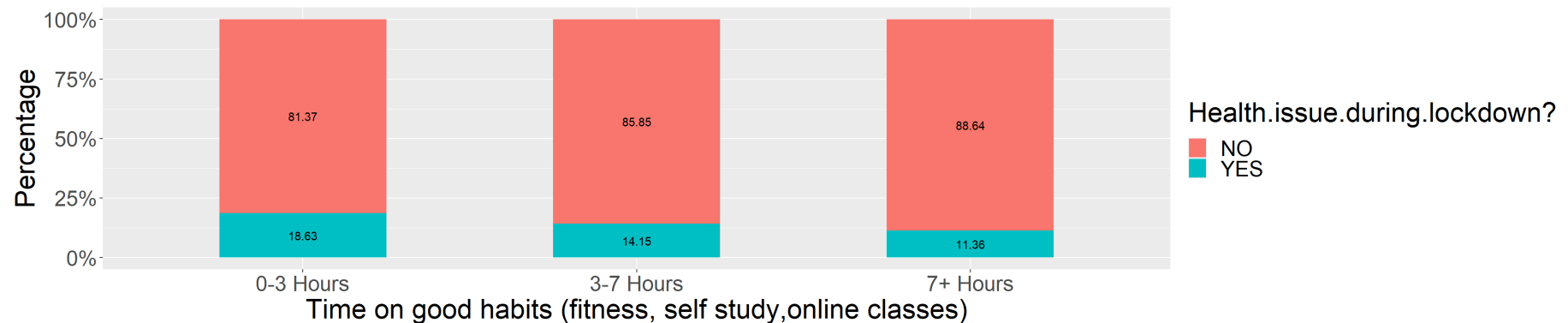
- Time spent on sleep
- Time spent on social media
- Number of meals per day
- Change in your weight

• Do you find yourself more connected with your family, close friends , relatives?

We find these features interesting, since according to researches these are the quite 'obvious' symptoms to evaluate whether the person is physically healthy (According to experts), or not, and whether the person's developing mental issues.

3. Preliminary results

Below we can see 2 graphs that indicate mental health by features that we checked

Description: we take the total amount of time that spent on good habits(fitness,self study and online class) and subtract the bad habits(time spent on social media and TV)

Prediction model

below you can see a prediction model that predicts by the following features: Time.spent.on.fitness, Time.spent.on.sleep, Time.spent.on.social.media, Time.spent.on.TV, Number.of.meals.per.day. weather he develop mental health issue and it represented by confusion matrix.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  NO YES
##        NO  388  20
##        YES  56   8
##
##               Accuracy : 0.839
##                 95% CI : (0.8027, 0.871)
##    No Information Rate : 0.9407
##    P-Value [Acc > NIR] : 1
##
##                  Kappa : 0.0996
##
##  Mcnemar's Test P-Value : 5.95e-05
##
##            Sensitivity : 0.8739
##            Specificity : 0.2857
##         Pos Pred Value : 0.9510
##         Neg Pred Value : 0.1250
##             Prevalence : 0.9407
##         Detection Rate : 0.8220
##   Detection Prevalence : 0.8644
##      Balanced Accuracy : 0.5798
##
##       'Positive' Class : NO
##
```

4. Data analysis plan

The limitations of our approach are: 1) The dataset checked students at specific place across Delhi National Capital Region(India) and it doesn't contain data of other countries during the lockdown. 2) If the dataset had more features that we can analyze, we could find more causes of mental health issues and get better understating and more accurate results of the consequences of one feature to an other(like sleeping and weight)

Future work: If had much more time we would done prediction models that predict student mental health by his answers to the features weather he will develop mental health issues or not in the future and try to prevent it before he will develop it, in addition of that we would check for more researchs that indicate our finding and new findings that reflect in our dataset.

Appendix

Signs and symptoms for poor mental health condition, presented by the American Psychiatric Association. These were used in order to obtain better understanding of which features should be further researched. https://psychiatry.org/patients-families/warning-signs-of-mental-illness (https://psychiatry.org/patients-families/warning-signs-of-mental-illness)

Data README

```
# SISE2601 Project data description
================
Team name: L


This Markdown file describes the data folder structure and organization ...


Features:


ID
Region of residence        <str> - checks weather the student is in Delhi-NCR or not
Age of Subject             <int> - the ages of the students between 7-59 years old
Time spent on Online Class <int> - checks student's time spent on Online Class, values between 0-10
Rating of Online Class experience  <str> - checks how student rated the online class by words: poor, avargae and  etc.
Medium for online class      <str> - shows how the students watched the online class with few catgories: Laptop/Desktop, smar
tphone and etc
Time spent on self study     <int> - checks how much time students spend their time of self study
Time spent on fitness        <int> - checks how much time students spend their time of fitness
Time spent on sleep          <int> - checks how much time students spend their time of sleeping
Time spent on social media   <int> - checks how much time students spend their time of social media
Prefered social media platform      <str> - shows students prefered social media platform: facebook, whatapp and etc.
Time spent on TV             <int> - checks how much time students spend their time on TV
Number of meals per day      <int> - checks how many meals students ate during the lockdown
Change in your weight        <str> - checks how student rated the weight of himself by words: incresing, decreased, Remain Co
nstant
Health issue during lockdown      <str> - checks if a student has health issue during the lockdown by saying yes/no
Stress busters               <str> - shows what are the stress busters according to students
Time utilized                <str> - shows if the time utilized during the lockdown by yes/no
Do you find yourself more connected with your family, close friends , relatives  ? <str> - shows if the lockdown got close b
etween family, close friends , relatives by saying yes/no
What you miss the most?      <str> - shows what the student missed the most during the lockdown: School/college , travlling a
nd etc.
```

Source code

```r
library(knitr)
library(tidyverse)
library(broom)
library(htmltools)
library(readr)
library(plyr)
library(dplyr)
library(caret)
library(caTools)
library(kernlab)
library(rpart)

opts_chunk$set(echo=FALSE) # hide source code in the document
df <- read.csv("C:\\Users\\Administrator\\Desktop\\data\\COVID-19 Survey Student Responses.csv")
#glimpse(df)


df$good_time <- rowSums(df[ , c(4,7,8)], na.rm=TRUE)
df <- df %>% mutate(group_good_time =
                    case_when(good_time >= 0 & good_time <= 3 ~ "0-3 Hours",
                              good_time > 3 & good_time <=  7~ "3-7 Hours",
                              good_time > 7 ~ "7+ Hours")
)

df$group_good_time <- factor(df$group_good_time, levels=c("0-3 Hours","3-7 Hours","7+ Hours"))

df <- df %>% mutate(hour_sleep =
                    case_when(Age.of.Subject >= 0 & Age.of.Subject < 15 ~ "0 - 15 Y.O.",
                              Age.of.Subject >= 15 & Age.of.Subject < 30 ~ "15 - 30 Y.O.",
                              Age.of.Subject >= 30 ~ "31+ Y.O.")
)


df$hour_sleep <- factor(df$hour_sleep, levels=c("0 - 15 Y.O.","15 - 30 Y.O.","31+ Y.O."))

ggplot(df, aes(x=group_good_time,fill = Health.issue.during.lockdown)) +

  geom_bar(position="fill",width = 0.5) +
```

```
  geom_text( aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)]*100, digits=4)), stat="count",
 position=position_fill(vjust=0.5)) +
  labs(y="Percentage",x="Time on good habits (fitness, self study,online classes)",fill="Health.issue.during.lockdown?")  +
  theme(text = element_text(size=25),plot.caption = element_text(hjust = 0, face = "bold"))+
  scale_y_continuous(labels = scales::percent_format())



df$Time.spent.on.TV[df$Time.spent.on.TV == "n" | df$Time.spent.on.TV == "N" | df$Time.spent.on.TV == "No tv" | df$Time.spen
t.on.TV == "" | df$Time.spent.on.TV == " " ] <- 0



df$Time.spent.on.TV <- as.double(df$Time.spent.on.TV)

df$good_time <- rowSums(df[ , c(4,7,8)], na.rm=TRUE)
df$screen_time <- rowSums(df[ , c(12,10)], na.rm=TRUE)
df$total <- df$good_time - df$screen_time

df7 <- df %>% mutate(group_screen =
                  case_when(total < -4  ~ "x<-4",
                            total >= -4 & total <= 3 ~ "-4<=x<=3",
                            total > 3 ~ "x>3")

)
df7$group_screen <- factor(df7$group_screen, levels=c("x<-4","-4<=x<=3","x>3"))


ggplot(df7, aes(x=group_screen,fill = Health.issue.during.lockdown)) +
  geom_bar(position="fill",width = 0.5)+
  geom_text( aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)]*100, digits=3)), stat="count",
 position=position_fill(vjust=0.5)) +
  labs(y="Percentage",x="Hours of good habbit - =Hours of bad habbit",fill="Health issue during lockdown", caption = "Descri
ption: we take the total amount of time that spent on \ngood habits(fitness,self study and online class) and subtract the ba
d habits(time spent on social media and TV)")  +
 theme(text = element_text(size=25),plot.caption = element_text(hjust = 0, face = "bold"))+
  scale_y_continuous(labels = scales::percent_format())

#options(width = 10)
```

```r
df <- read.csv("C:\\Users\\Administrator\\Desktop\\COVID-19 Survey Student Responses.csv")
df1 <- select(df, Time.spent.on.fitness, Time.spent.on.sleep, Time.spent.on.social.media, Time.spent.on.TV, Number.of.meals.
per.day, Change.in.your.weight, Time.utilized,Age.of.Subject, Health.issue.during.lockdown)
set.seed(100)
colms = c(1,2,3,4,5,6,7,8,9)
df1[,colms] <- lapply(df1[,colms], factor)
trainindex <- createDataPartition(df1$Health.issue.during.lockdown, p=0.6, list = FALSE)
train_work <- df1[trainindex,]
test_work <- df1[-trainindex,]

Model1 <- train(Health.issue.during.lockdown ~ ., data = train_work, method = 'svmPoly', na.action = na.omit, preProcess = c
("scale", "center"), trControl = trainControl(method = 'none'), tuneGrid = data.frame(degree=1, scale=1, C=1), metric = 'Acc
uracy')

model = rpart(Health.issue.during.lockdown~., data = train_work, method = 'class')
test_work$prediction <- predict(Model1, test_work)
test_work$factor_pred <- as.factor(test_work$prediction)
test_work$factor_truth <- as.factor(test_work$Health.issue.during.lockdown)
predict_model <- predict(model, test_work, type = "class")

#precision_val <- posPredValue(test_work$factor_truth, test_work$factor_pred)
confusion_mat = confusionMatrix(test_work$Health.issue.during.lockdown, predict_model)
#posPredValue(test_work$Health.issue.during.lockdown, predict_model)


#con_mat <- confusionMatrix(test_df$prediction, as.factor(test_df$Health.issue.during.lockdown))
#confusion_mat = table(test_work$Health.issue.during.lockdown, predict_model)
#recall_val <- sensitivity(test_df$factor_truth, test_df$factor_pred)
print(confusion_mat)
#precision_val
#recall_val
cat(readLines('../data/README.md'), sep = '\n')
```