Lesson 3 Practice Hands-On

Directions

Recoding Hands-On

This Hands-On will **not** be graded, but you are encouraged to complete it. The best way to become a data scientist is to practice.

Caution!

Do not submit your project until you have completed all requirements, as you will not be able to resubmit.

You are working for a global chocolate company, and they've collected **data on worldwide eating habits**. Their eventual goal is to determine the demographics for chocolate-eaters worldwide. Which countries are most likely to consume chocolate? Which gender, and which age group? Someone else will run these analyses, but it is your job to wrangle and recode the data in preparation.

Part 1: Recoding in Python

Please perform the following tasks in Python:

- Recode `Activity` into a new variable. Zeros should be not eating chocolate, and 1s should be eating chocolate.
- Recode `Frequency` from text to numbers in the same variable. The value zero should be the lowest frequency.
- Recode `Sex` from text to numbers in the same variable.
- Dummy code the `Age` group variable.

Part 2: Recoding in R

Please perform the following tasks in R:

- Recode `Activity` into a new variable called `JunkFood`. Anything that you would consider junk food, recode as a 1. Everything else should be recoded as a zero.
- Recode `Sex` from text to numbers in the same variable

Please comment your code and include both your Python and R files.

Caution!

Be sure to zip and submit your entire directory when finished!

Solution

Part 1: Recoding in Python

```
import pandas as pd

EatingHabits = pd.read_csv('C:/Users/meredith.dodd/Documents/New
Curriculum/Recoding/Eating_Habits.csv')
EatingHabits.head()

#Recode Activity into a New Variable

EatingHabits.Activity.value_counts()

def activity (series):
    if series == "Eating fruit" :
        return 0
    if series == "Drinking soft drinks, cola or other drinks
with sugar" :
        return 0
    if series == "Drinking coffee" :
        return 0
    if series == "Eating french fries" :
        return 0
    if series == "Eating hamburgers, hot dogs or sausages" :
        return 0
    if series == "Eating candy, chocolate bars" :
        return 1
    if series == "Eating whole wheat or rye bread" :
        return 0
    if series == "Eating raw vegetables" :
        return 0
    if series == "Eating potato chips, crisps" :
        return 0
    if series == "Drinking whole milk" :
        return 0
    if series == "Drinking low fat milk" :
```

```
        return 0
    if series == "Eating peanuts" :
        return 0

EatingHabits['ChocolateYN'] =
EatingHabits['Activity'].apply(activity)
EatingHabits.head()

# Recode Frequency and Sex from text to numbers in the same
variable

EatingHabits.Frequency.value_counts()
EatingHabits.Sex.value_counts()

cleanup = {"Frequency" : {"Never" : 0, "Seldom" : 1, "At least
once a week" : 2, "Once a day" : 3, "More than once a day" : 4},
"Sex" : {"Females" : 0, "Males" : 1}}
EatingHabits.replace(cleanup, inplace=True)
EatingHabits.head()

#Dummy code the Age group variable

AgeDummy = pd.get_dummies(EatingHabits['Age group'],
drop_first=True)
AgeDummy

EatingHabits2 = pd.concat([EatingHabits, AgeDummy], axis=1)
EatingHabits2.head()
```

Part 2: Recoding in R

To recode activity into a new variable called JunkFood

```
Eating_Habits$JunkFood < NA
Eating_Habits$JunkFood[Eating_Habits$Activity=='Eating fruit']
<- 0
Eating_Habits$JunkFood[Eating_Habits$Activity=='Eating raw
vegetables'] <- 0
```

```
Eating_Habits$JunkFood[Eating_Habits$Activity=='Eating candy,
chocolate bars'] <- 1
Eating_Habits$JunkFood[Eating_Habits$Activity=='Eating potato
chips, crisps'] <- 1
Eating_Habits$JunkFood[Eating_Habits$Activity=='Eating french
fries'] <- 1
Eating_Habits$JunkFood[Eating_Habits$Activity=='Eating
hamburgers, hot dogs or sausages'] <- 1
Eating_Habits$JunkFood[Eating_Habits$Activity=='Eating peanuts']
<- 0
Eating_Habits$JunkFood[Eating_Habits$Activity=='Eating whole
wheat or rye bread'] <- 0
Eating_Habits$JunkFood[Eating_Habits$Activity=='Drinking soft
drinks, cola or other drinks with sugar'] <- 1
Eating_Habits$JunkFood[Eating_Habits$Activity=='Drinking
coffee'] <- 0
Eating_Habits$JunkFood[Eating_Habits$Activity=='Eating Fruit']
<- 0
```

To recode sex from text to numbers in the same variable:

```
Eating_Habits$Sex[Eating_Habits$Sex=='Males'] <- 0
Eating_Habits$Sex[Eating_Habits$Sex=='Females'] <- 1
```

To dummy code the frequency variable:

```
library("psych")

Eating_Habits1 <- dummy.code(Eating_Habits$Frequency)
Eating_Habits2 <- data.frame(Eating_Habits, Eating_Habits1)
```