RWorksheet_Tubat#3

2023-10-11

```
#1a. Write the codes
respondents <- seq(1,20)
sex <- c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,2,1,2)
fathersOccupation <- c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,3,1,3,1,2,1)
personsAtHome <- c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
siblingsAtSchool <- c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
typesOfHouses <- c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)
df <- data.frame(respondents,sex,fathersOccupation,personsAtHome,siblingsAtSchool,typesOfHouses)
df
```

```
##    respondents sex fathersOccupation personsAtHome siblingsAtSchool
## 1            1   2                 1             5                6
## 2            2   2                 3             7                4
## 3            3   1                 3             3                4
## 4            4   2                 3             8                1
## 5            5   2                 1             5                2
## 6            6   2                 2             9                1
## 7            7   2                 3             6                5
## 8            8   2                 1             7                3
## 9            9   2                 1             8                1
## 10          10   2                 1             4                2
## 11          11   1                 3             7                3
## 12          12   2                 2             5                2
## 13          13   2                 1             4                5
## 14          14   2                 3             7                5
## 15          15   2                 3             8                2
## 16          16   2                 1             8                1
## 17          17   2                 3             3                2
## 18          18   2                 1            11                5
## 19          19   1                 2             7                3
## 20          20   2                 1             6                2
##    typesOfHouses
## 1              1
## 2              2
## 3              3
## 4              1
## 5              1
## 6              3
## 7              3
## 8              1
## 9              2
## 10             3
## 11             2
## 12             3
## 13             2
```

```
## 14             2
## 15             3
## 16             3
## 17             3
## 18             3
## 19             3
## 20             2
```

```
#1b. Describe the data. Get the structure or the summary of the data
explainData <- summary(df)
explainData
```

```
##   respondents         sex       fathersOccupation personsAtHome
##  Min.   : 1.00   Min.   :1.00   Min.   :1.00      Min.   : 3.0
##  1st Qu.: 5.75   1st Qu.:2.00   1st Qu.:1.00      1st Qu.: 5.0
##  Median :10.50   Median :2.00   Median :2.00      Median : 7.0
##  Mean   :10.50   Mean   :1.85   Mean   :1.95      Mean   : 6.4
##  3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.00      3rd Qu.: 8.0
##  Max.   :20.00   Max.   :2.00   Max.   :3.00      Max.   :11.0
##  siblingsAtSchool typesOfHouses
##  Min.   :1.00     Min.   :1.0
##  1st Qu.:2.00     1st Qu.:2.0
##  Median :2.50     Median :2.5
##  Mean   :2.95     Mean   :2.3
##  3rd Qu.:4.25     3rd Qu.:3.0
##  Max.   :6.00     Max.   :3.0
```

```
#1c. Is the mean number of siblings attending is 5?
#The mean number of siblings attending is 2.95 not 5.
meanNumber <- mean(siblingsAtSchool)
meanNumber
```

```
## [1] 2.95
```

```
#1d. Extract the 1st two rows and then all the columns using the subsetting functions.
#Write the codes and its output.
df[1:2,]
```

```
##   respondents sex fathersOccupation personsAtHome siblingsAtSchool
## 1           1   2                 1             5                6
## 2           2   2                 3             7                4
##   typesOfHouses
## 1             1
## 2             2
```

```
#1e. Extract 3rd and 5th row with 2nd and 4th column. Write the codes and its result.
df[c(3,5),c(2,4)]
```

```
##   sex personsAtHome
## 3   1             3
## 5   2             5
```

```
#1f. Select the variable types of houses then store the vector that results as types_houses.
#Write the codes.

types_houses <- df[,6]
types_houses
```

```
##  [1] 1 2 3 1 1 3 3 1 2 3 2 3 2 2 3 3 3 3 3 2
```

```
#1Select only all Males respondent that their father occupation was farmer.
#Write the codes and its output.
combinedData <- subset(df, sex == 1 & fathersOccupation == 1)
combinedData
```

```
## [1] respondents       sex               fathersOccupation personsAtHome
## [5] siblingsAtSchool  typesOfHouses
## <0 rows> (or 0-length row.names)
```

```
#1h. Select only all females respondent that have greater than or equal to 5 number of
#siblings attending school. Write the codes and its outputs.
combinedData2 <- subset(df, sex == 2 & siblingsAtSchool >= 5)
combinedData2
```

```
##    respondents sex fathersOccupation personsAtHome siblingsAtSchool
## 1            1   2                 1             5                6
## 7            7   2                 3             6                5
## 13          13   2                 1             4                5
## 14          14   2                 3             7                5
## 18          18   2                 1            11                5
##    typesOfHouses
## 1              1
## 7              3
## 13             2
## 14             2
## 18             3
```

```
df = data.frame(Ints=integer(),
  Doubles=double(), Characters=character(),
  Logicals=logical(),
  Factors=factor(),
  stringsAsFactors=FALSE)

print("Structure of the empty dataframe:")
```

```
## [1] "Structure of the empty dataframe:"
```

```
print(str(df))
```

```
## 'data.frame':    0 obs. of  5 variables:
##  $ Ints      : int
##  $ Doubles   : num
##  $ Characters: chr
##  $ Logicals  : logi
##  $ Factors   : Factor w/ 0 levels:
## NULL
```

```
#The empty dataframe has different kinds of data types namely: int, doubles, characters, logical, and f
```

```
#3a. Import the csv file into the R environment. Write the codes
houseHoldData <- read.csv("/cloud/project/worksheet#3/HouseholdData.csv")
houseHoldData
```

```
##    Respondents    Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1            1   Male                 1              5                 2
## 2            2 Female                 2              7                 3
## 3            3 Female                 3              3                 0
```

```
## 4            4   Male                 3               8                 5
## 5            5   Male                 1               6                 2
## 6            6 Female                 2               4                 3
## 7            7 Female                 2               4                 1
## 8            8   Male                 3               2                 2
## 9            9 Female                 1              11                 6
## 10          10   Male                 3               6                 2
##    Types.of.Houses
## 1             Wood
## 2         Congrete
## 3         Congrete
## 4             Wood
## 5    Semi-congrete
## 6    Semi-congrete
## 7             Wood
## 8    Semi-congrete
## 9    Semi-congrete
## 10        Congrete
```

```
#3b. Convert the Sex into factor using factor() function and change it into integer.[Legend:
# Male = 1 and Female = 2]. Write the R codes and its output.

houseHoldData$Sex <- factor(houseHoldData$Sex, levels = c("Male" , "Female"), label = c(1,2))
houseHoldData
```

```
##    Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1            1   1                  1               5                 2
## 2            2   2                  2               7                 3
## 3            3   2                  3               3                 0
## 4            4   1                  3               8                 5
## 5            5   1                  1               6                 2
## 6            6   2                  2               4                 3
## 7            7   2                  2               4                 1
## 8            8   1                  3               2                 2
## 9            9   2                  1              11                 6
## 10          10   1                  3               6                 2
##    Types.of.Houses
## 1             Wood
## 2         Congrete
## 3         Congrete
## 4             Wood
## 5    Semi-congrete
## 6    Semi-congrete
## 7             Wood
## 8    Semi-congrete
## 9    Semi-congrete
## 10        Congrete
```

```
#3c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood
# = 1; Congrete = 2; Semi-Congrete = 3]. Write the R codes and its output.
houseHoldData$Types.of.Houses <- factor(houseHoldData$Types.of.Houses, levels = c("Wood", "Congrete", "
houseHoldData
```

```
##    Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1            1   1                  1               5                 2
## 2            2   2                  2               7                 3
```

```
## 3            3   2                 3               3                 0
## 4            4   1                 3               8                 5
## 5            5   1                 1               6                 2
## 6            6   2                 2               4                 3
## 7            7   2                 2               4                 1
## 8            8   1                 3               2                 2
## 9            9   2                 1              11                 6
## 10          10   1                 3               6                 2
##    Types.of.Houses
## 1                1
## 2                2
## 3                2
## 4                1
## 5                3
## 6                3
## 7                1
## 8                3
## 9                3
## 10               2
```

```
#3d. On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3. What
#is the R code and its output?
houseHoldData$Fathers.Occupation <- factor(houseHoldData$Fathers.Occupation, levels = c(1,2,3), label =
houseHoldData
```

```
##    Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 1            1   1             Farmer               5                 2
## 2            2   2             Driver               7                 3
## 3            3   2             Others               3                 0
## 4            4   1             Others               8                 5
## 5            5   1             Farmer               6                 2
## 6            6   2             Driver               4                 3
## 7            7   2             Driver               4                 1
## 8            8   1             Others               2                 2
## 9            9   2             Farmer              11                 6
## 10          10   1             Others               6                 2
##    Types.of.Houses
## 1                1
## 2                2
## 3                2
## 4                1
## 5                3
## 6                3
## 7                1
## 8                3
## 9                3
## 10               2
```

```
#3e. Select only all females respondent that has a father whose occupation is driver. Write
#the codes and its output.
selectFemaleDriver <- subset(houseHoldData, Sex == 2 & Fathers.Occupation == "Driver")
selectFemaleDriver
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 2           2   2             Driver               7                 3
## 6           6   2             Driver               4                 3
```

```
## 7           7   2             Driver                4                   1
##   Types.of.Houses
## 2               2
## 6               3
## 7               1
```

```
#3f. Select the respondents that have greater than or equal to 5 number of siblings attending
#school. Write the codes and its output.
respondent <- subset(houseHoldData, houseHoldData$Siblings.at.School >= 5)
respondent
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 4           4   1             Others               8                  5
## 9           9   2             Farmer              11                  6
##   Types.of.Houses
## 4               1
## 9               3
```

```
knitr::include_graphics("/cloud/project/worksheet#3/twitter.png")
```

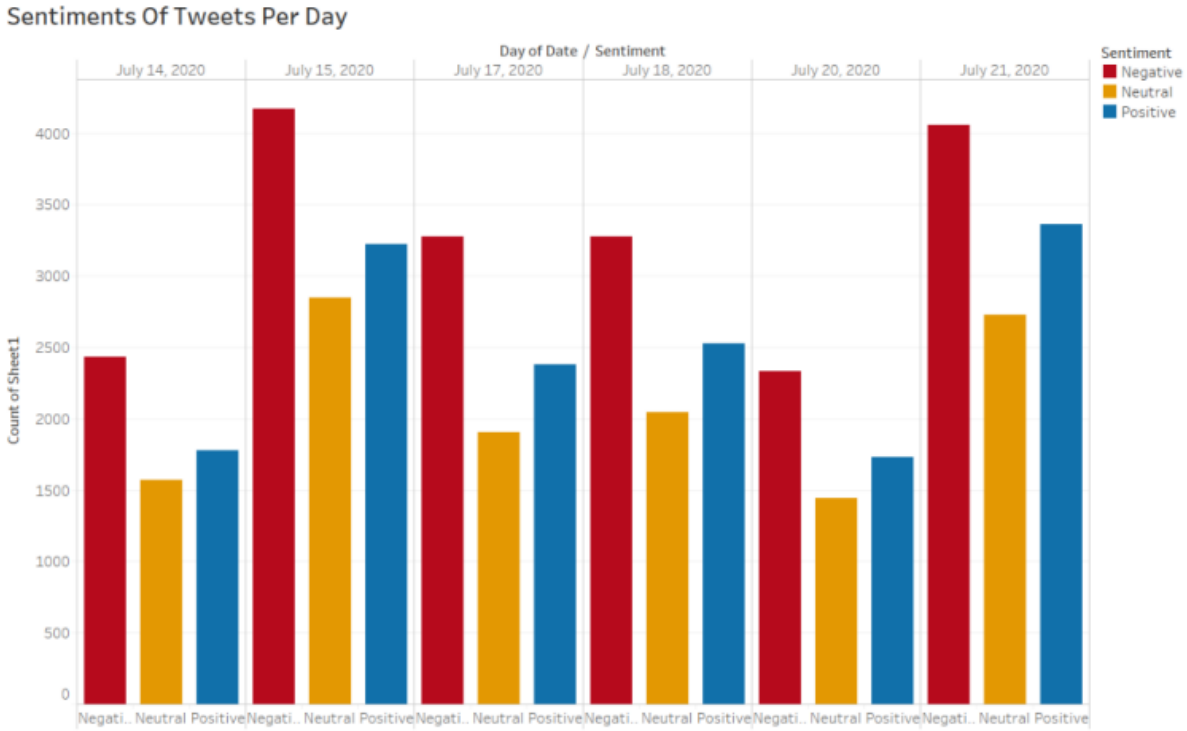

Figure 3: Figure 2: Sentiment Analysis

```
#This is a bar graph that has a title of Sentiments Of Tweets Per day the Y axis has the values from 0

#Most of the sentiments of tweets are negative and the highest number of negative sentiments are on Jul

#This data is a sentiment analysis and it shows that most of the tweets in twitter are negative.
```