RWorksheet_Gonzaga#3b

SHAMEL GONZAGA

2024-10-02

1

A.

```
df <- data.frame(
  Respondents = 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
```

```
##    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
```

B.

```
str(df)
```

```
## 'data.frame':    20 obs. of  6 variables:
##  $ Respondents      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Sex              : num  2 2 1 2 2 2 2 2 2 2 ...
##  $ FathersOccupation: num  1 3 3 3 1 2 3 1 1 1 ...
##  $ PersonsAtHome    : num  5 7 3 8 5 9 6 7 8 4 ...
##  $ SiblingsAtSchool : num  6 4 4 1 2 1 5 3 1 2 ...
##  $ TypesOfHouses    : num  1 2 3 1 1 3 3 1 2 3 ...
```

The structure represents quantity of objects and variables in the data frame. It provides the initial few contents of the dataframe along with the data type of every column.

C.

```
mean(df$SiblingsAtSchool)
```

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

D.

```r
subset<- df[1:2, ]
subset
```

```
##   Respondents Sex FathersOccupation PersonsAtHome SiblingsAtSchool
## 1           1   2                 1             5                6
## 2           2   2                 3             7                4
##   TypesOfHouses
## 1             1
## 2             2
```

E.

```r
subSetOne <- df[c(3, 5), c(2, 4)]
subSetOne
```

```
##   Sex PersonsAtHome
## 3   1             3
## 5   2             5
```

F.

```r
types.houses <- df$TypesOfHouses
```

G.

```r
maleFarmers <- subset(df, Sex == 1 & FathersOccupation == 1)
maleFarmers
```

```
## [1] Respondents       Sex               FathersOccupation PersonsAtHome
## [5] SiblingsAtSchool  TypesOfHouses
## <0 rows> (or 0-length row.names)
```

H.

```r
femaleSiblings <- subset(df, Sex == 2 & SiblingsAtSchool >= 5)
femaleSiblings
```

```
##    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
```

2.

```
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
```

A.

#Looking at the result it will shows the structure of the empty data frame. As depicted therein, it has 0 observations and 5 variables. It also displays different data types with no data yet

3.

A.

```
HouseHold <- read.csv("HouseholdData.csv")
HouseHold
```

```
##    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
```

B.

```
HouseHold$Sex <- as.integer(factor(HouseHold$Sex,
                  levels = c("Male", "Female"),
                  labels = c(1, 2)))
HouseHold
```

```
##    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
```

C.

```
HouseHold$Types.of.Houses <- as.integer(factor(HouseHold$Types.of.Houses, levels = c("Wood", "Congrete"
HouseHold
```

```
##    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
```

D.

```
HouseHold$Fathers_Occupation <-  as.character(factor(HouseHold$Fathers.Occupation, levels = c(1, 2, 3),
HouseHold$Fathers_Occupation
```

```
##  [1] "Farmer" "Driver" "Others" "Others" "Farmer" "Driver" "Driver" "Others"
##  [9] "Farmer" "Others"
```

E.

```
femaleDriver <- subset(HouseHold, Sex == 2 & Fathers_Occupation == "Driver")
femaleDriver
```

```
##   Respondents Sex Fathers.Occupation Persons.at.Home Siblings.at.School
## 2           2   2                  2               7                  3
## 6           6   2                  2               4                  3
## 7           7   2                  2               4                  1
##   Types.of.Houses Fathers_Occupation
```

```
## 2               2             Driver
## 6               3             Driver
## 7               1             Driver
```

F.

```
SiblingSchool <- subset(HouseHold, Siblings.at.School >= 5)
SiblingSchool
```

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

4.

The graph analysis of the tweet reveals that the highest number of the tweets per day from July 14, 2020 to July 21, 2020 has shown negative. The second highest percentage is positive while the lowest percentage of sentiments of tweets per day is neutral.