Student Stress Factors Data Science Project

Berke Kaygısız

2024-12-17

# GANT CHART:

metin, ekran görüntüsü, diyagram, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

### 1-PROBLEM:

Stress is one of the most important factors that negatively affect individuals’ mental, physical and social health. High stress levels can cause poor performance in business, education and personal life, various health problems and sleep disorders. Individuals’ stress levels are affected by various psychological, physiological and social factors. One of the biggest problems of students is stress.

This data science project aims to analyze the effects of these factors (anxiety, depression, self-esteem, sleep quality, bullying, etc.) on students’ stress levels, reveal the causes of stress and provide insights into how students can better manage their stress levels in line with these reasons.

#### Import Data Set

library(readxl)  
Stress\_level\_data <- read\_excel("C:/Users/berke/Desktop/Stress\_level\_data.xlsx")  
View(Stress\_level\_data)

## 2-Exploratry Data Analysis

#### Overview of Dataset Structure and Summary Statistics

dim(Stress\_level\_data)

## [1] 1100 21

head(Stress\_level\_data)

## # A tibble: 6 × 21  
## anxiety\_level self\_esteem mental\_health\_history depression headache  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 14 20 0 11 2  
## 2 15 8 1 15 5  
## 3 12 18 1 14 2  
## 4 16 12 1 15 4  
## 5 16 28 0 7 2  
## 6 20 13 1 21 3  
## # ℹ 16 more variables: blood\_pressure <dbl>, sleep\_quality <dbl>,  
## # breathing\_problem <dbl>, noise\_level <dbl>, living\_conditions <dbl>,  
## # safety <dbl>, basic\_needs <dbl>, academic\_performance <dbl>,  
## # study\_load <dbl>, teacher\_student\_relationship <dbl>,  
## # future\_career\_concerns <dbl>, social\_support <dbl>, peer\_pressure <dbl>,  
## # extracurricular\_activities <dbl>, bullying <dbl>, stress\_level <dbl>

colnames(Stress\_level\_data)

## [1] "anxiety\_level" "self\_esteem"   
## [3] "mental\_health\_history" "depression"   
## [5] "headache" "blood\_pressure"   
## [7] "sleep\_quality" "breathing\_problem"   
## [9] "noise\_level" "living\_conditions"   
## [11] "safety" "basic\_needs"   
## [13] "academic\_performance" "study\_load"   
## [15] "teacher\_student\_relationship" "future\_career\_concerns"   
## [17] "social\_support" "peer\_pressure"   
## [19] "extracurricular\_activities" "bullying"   
## [21] "stress\_level"

summary(Stress\_level\_data)

## anxiety\_level self\_esteem mental\_health\_history depression   
## Min. : 0.00 Min. : 0.00 Min. :0.0000 Min. : 0.00   
## 1st Qu.: 6.00 1st Qu.:11.00 1st Qu.:0.0000 1st Qu.: 6.00   
## Median :11.00 Median :19.00 Median :0.0000 Median :12.00   
## Mean :11.06 Mean :17.78 Mean :0.4927 Mean :12.56   
## 3rd Qu.:16.00 3rd Qu.:26.00 3rd Qu.:1.0000 3rd Qu.:19.00   
## Max. :21.00 Max. :30.00 Max. :1.0000 Max. :27.00   
## headache blood\_pressure sleep\_quality breathing\_problem  
## Min. :0.000 Min. :1.000 Min. :0.00 Min. :0.000   
## 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.00 1st Qu.:2.000   
## Median :3.000 Median :2.000 Median :2.50 Median :3.000   
## Mean :2.508 Mean :2.182 Mean :2.66 Mean :2.754   
## 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:4.00 3rd Qu.:4.000   
## Max. :5.000 Max. :3.000 Max. :5.00 Max. :5.000   
## noise\_level living\_conditions safety basic\_needs   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000   
## Median :3.000 Median :2.000 Median :2.000 Median :3.000   
## Mean :2.649 Mean :2.518 Mean :2.737 Mean :2.773   
## 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000   
## academic\_performance study\_load teacher\_student\_relationship  
## Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000   
## Median :2.000 Median :2.000 Median :2.000   
## Mean :2.773 Mean :2.622 Mean :2.648   
## 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:4.000   
## Max. :5.000 Max. :5.000 Max. :5.000   
## future\_career\_concerns social\_support peer\_pressure   
## Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:2.000   
## Median :2.000 Median :2.000 Median :2.000   
## Mean :2.649 Mean :1.882 Mean :2.735   
## 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:4.000   
## Max. :5.000 Max. :3.000 Max. :5.000   
## extracurricular\_activities bullying stress\_level   
## Min. :0.000 Min. :0.000 Min. :0.0000   
## 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:0.0000   
## Median :2.500 Median :3.000 Median :1.0000   
## Mean :2.767 Mean :2.617 Mean :0.9964   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:2.0000   
## Max. :5.000 Max. :5.000 Max. :2.0000

colSums(is.na(Stress\_level\_data))

## anxiety\_level self\_esteem   
## 0 0   
## mental\_health\_history depression   
## 0 0   
## headache blood\_pressure   
## 0 0   
## sleep\_quality breathing\_problem   
## 0 0   
## noise\_level living\_conditions   
## 0 0   
## safety basic\_needs   
## 0 0   
## academic\_performance study\_load   
## 0 0   
## teacher\_student\_relationship future\_career\_concerns   
## 0 0   
## social\_support peer\_pressure   
## 0 0   
## extracurricular\_activities bullying   
## 0 0   
## stress\_level   
## 0

str(Stress\_level\_data)

## tibble [1,100 × 21] (S3: tbl\_df/tbl/data.frame)  
## $ anxiety\_level : num [1:1100] 14 15 12 16 16 20 4 17 13 6 ...  
## $ self\_esteem : num [1:1100] 20 8 18 12 28 13 26 3 22 8 ...  
## $ mental\_health\_history : num [1:1100] 0 1 1 1 0 1 0 1 1 0 ...  
## $ depression : num [1:1100] 11 15 14 15 7 21 6 22 12 27 ...  
## $ headache : num [1:1100] 2 5 2 4 2 3 1 4 3 4 ...  
## $ blood\_pressure : num [1:1100] 1 3 1 3 3 3 2 3 1 3 ...  
## $ sleep\_quality : num [1:1100] 2 1 2 1 5 1 4 1 2 1 ...  
## $ breathing\_problem : num [1:1100] 4 4 2 3 1 4 1 5 4 2 ...  
## $ noise\_level : num [1:1100] 2 3 2 4 3 3 1 3 3 0 ...  
## $ living\_conditions : num [1:1100] 3 1 2 2 2 2 4 1 3 5 ...  
## $ safety : num [1:1100] 3 2 3 2 4 2 4 1 3 2 ...  
## $ basic\_needs : num [1:1100] 2 2 2 2 3 1 4 1 3 2 ...  
## $ academic\_performance : num [1:1100] 3 1 2 2 4 2 5 1 3 2 ...  
## $ study\_load : num [1:1100] 2 4 3 4 3 5 1 3 3 2 ...  
## $ teacher\_student\_relationship: num [1:1100] 3 1 3 1 1 2 4 2 2 1 ...  
## $ future\_career\_concerns : num [1:1100] 3 5 2 4 2 5 1 4 3 5 ...  
## $ social\_support : num [1:1100] 2 1 2 1 1 1 3 1 3 1 ...  
## $ peer\_pressure : num [1:1100] 3 4 3 4 5 4 2 4 3 5 ...  
## $ extracurricular\_activities : num [1:1100] 3 5 2 4 0 4 2 4 2 3 ...  
## $ bullying : num [1:1100] 2 5 2 5 5 5 1 5 2 4 ...  
## $ stress\_level : num [1:1100] 1 2 1 2 1 2 0 2 1 1 ...

colSums(is.na(Stress\_level\_data))

## anxiety\_level self\_esteem   
## 0 0   
## mental\_health\_history depression   
## 0 0   
## headache blood\_pressure   
## 0 0   
## sleep\_quality breathing\_problem   
## 0 0   
## noise\_level living\_conditions   
## 0 0   
## safety basic\_needs   
## 0 0   
## academic\_performance study\_load   
## 0 0   
## teacher\_student\_relationship future\_career\_concerns   
## 0 0   
## social\_support peer\_pressure   
## 0 0   
## extracurricular\_activities bullying   
## 0 0   
## stress\_level   
## 0

str(Stress\_level\_data)

## tibble [1,100 × 21] (S3: tbl\_df/tbl/data.frame)  
## $ anxiety\_level : num [1:1100] 14 15 12 16 16 20 4 17 13 6 ...  
## $ self\_esteem : num [1:1100] 20 8 18 12 28 13 26 3 22 8 ...  
## $ mental\_health\_history : num [1:1100] 0 1 1 1 0 1 0 1 1 0 ...  
## $ depression : num [1:1100] 11 15 14 15 7 21 6 22 12 27 ...  
## $ headache : num [1:1100] 2 5 2 4 2 3 1 4 3 4 ...  
## $ blood\_pressure : num [1:1100] 1 3 1 3 3 3 2 3 1 3 ...  
## $ sleep\_quality : num [1:1100] 2 1 2 1 5 1 4 1 2 1 ...  
## $ breathing\_problem : num [1:1100] 4 4 2 3 1 4 1 5 4 2 ...  
## $ noise\_level : num [1:1100] 2 3 2 4 3 3 1 3 3 0 ...  
## $ living\_conditions : num [1:1100] 3 1 2 2 2 2 4 1 3 5 ...  
## $ safety : num [1:1100] 3 2 3 2 4 2 4 1 3 2 ...  
## $ basic\_needs : num [1:1100] 2 2 2 2 3 1 4 1 3 2 ...  
## $ academic\_performance : num [1:1100] 3 1 2 2 4 2 5 1 3 2 ...  
## $ study\_load : num [1:1100] 2 4 3 4 3 5 1 3 3 2 ...  
## $ teacher\_student\_relationship: num [1:1100] 3 1 3 1 1 2 4 2 2 1 ...  
## $ future\_career\_concerns : num [1:1100] 3 5 2 4 2 5 1 4 3 5 ...  
## $ social\_support : num [1:1100] 2 1 2 1 1 1 3 1 3 1 ...  
## $ peer\_pressure : num [1:1100] 3 4 3 4 5 4 2 4 3 5 ...  
## $ extracurricular\_activities : num [1:1100] 3 5 2 4 0 4 2 4 2 3 ...  
## $ bullying : num [1:1100] 2 5 2 5 5 5 1 5 2 4 ...  
## $ stress\_level : num [1:1100] 1 2 1 2 1 2 0 2 1 1 ...

The data set consists of 1100 observations and 21 variables.

The target variable of the data set is stress\_level.

There are no missing values in the data set.

When we look at the summary statistics, the following comments can be made. Such comments can be diversified.

Since the average value of the anxiety\_level variable is 11.06, we can say that the average anxiety level of the students is 11.06.

Since the average value of the self\_esteem variable is 17.78, we can say that the average self-confidence level of the students is 17.78.

Since the average value of the depression variable is 12.56, we can say that the average depression level of the students is 12.56.

Since the average value of the sleep\_quality variable is 2.66, we can say that the students’ sleep quality is at a moderate level.

The variables that are thought to have an effect on stress can be grouped as follows:

Psychological Factors => ‘anxiety\_level’, ‘self\_esteem’, ‘mental\_health\_history’, ‘depression’, Physiological Factors => ‘headache’, ‘blood\_pressure’, ‘sleep\_quality’, ‘breathing\_problem Environmental Factors => ’noise\_level’, ‘living\_conditions’, ‘safety’, ‘basic\_needs’, Academic Factors => ‘academic\_performance’, ‘study\_load’, ‘teacher\_student\_relationship’, ‘future\_career\_concerns’, Social Factor => ‘social\_support’, ‘peer\_pressure’, ‘extracurricular\_activities’, ‘bullying’

#### Analysis of Variables:

anxiety\_level: Shows the student’s anxiety level. Higher values represent higher anxiety levels.

self\_esteem: Shows the student’s self-confidence level. Higher values mean higher self-confidence.

mental\_health\_history: Shows whether the student has had mental health problems in the past. (A binary variable, Yes-No.)

depression: Shows the student’s depression level. Higher values indicate more serious depression symptoms.

headache: Shows the student’s headache severity. Sortable categorical variable.

blood\_pressure: Shows the student’s blood pressure status. Sortable categorical variable.

sleep\_quality: Shows the student’s sleep quality level. Sortable categorical variable.

breathing\_problem: Shows the student’s breathing problem level. Sortable categorical variable.

noise\_level: Shows the noise level in the student’s environment. Sortable categorical variable.

living\_conditions: Shows the quality of the student’s living conditions. Sortable categorical variable.

safety: Indicates the student’s perception of safety. It is an orderable categorical variable.

basic\_needs: Indicates the level of student’s basic needs being met. It is an orderable categorical variable.

academic\_performance: Indicates the student’s academic success level. It is an orderable categorical variable.

study\_load: Indicates the student’s course load or work intensity. It is an orderable categorical variable.

teacher\_student\_relationship: Indicates the quality of the teacher-student relationship. It is an orderable categorical variable.

future\_career\_concerns: Indicates the student’s future career concerns. It is an orderable categorical variable.

social\_support: Indicates the student’s social support level. It is an orderable categorical variable.

peer\_pressure: Indicates the student’s level of peer pressure. It is an orderable categorical variable.

extracurricular\_activities: Indicates the number of extracurricular activities the student participates in. It is an orderable categorical variable.

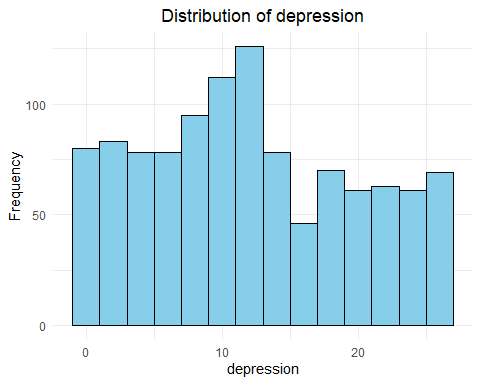
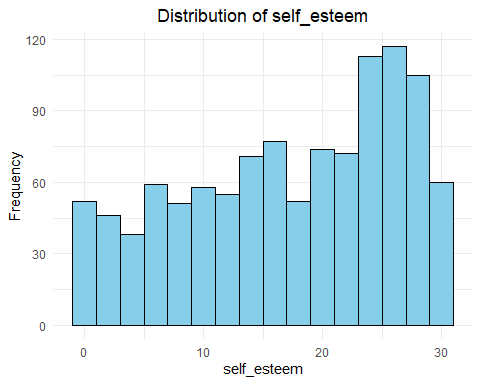
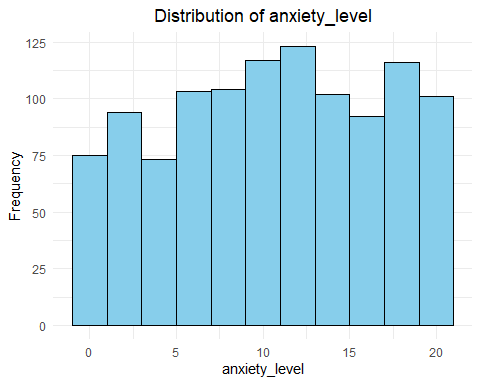
bullying: Indicates the student’s level of bullying. It is an orderable categorical variable.

stress\_level: Shows the stress level of the student. It is a sortable categorical variable.

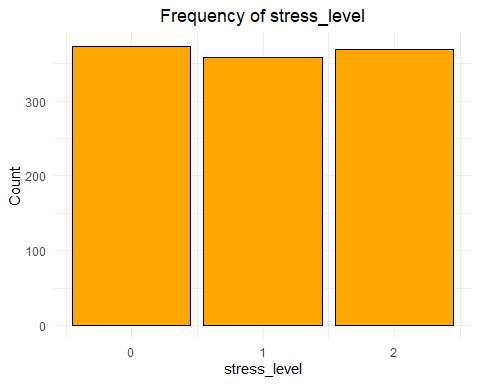
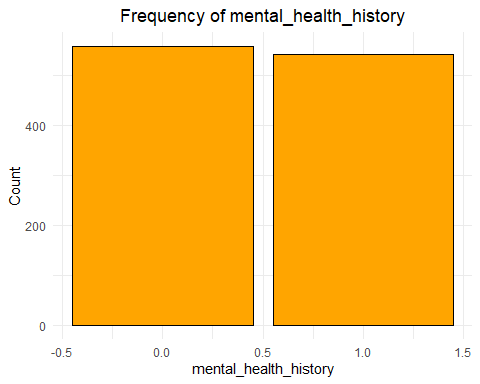
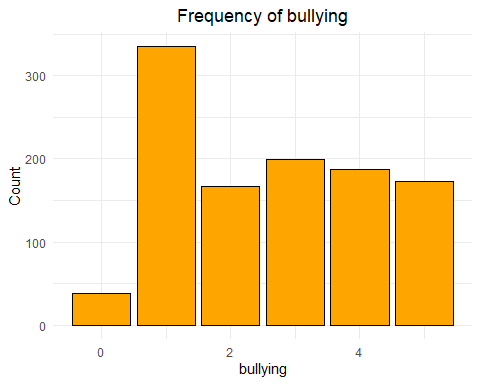
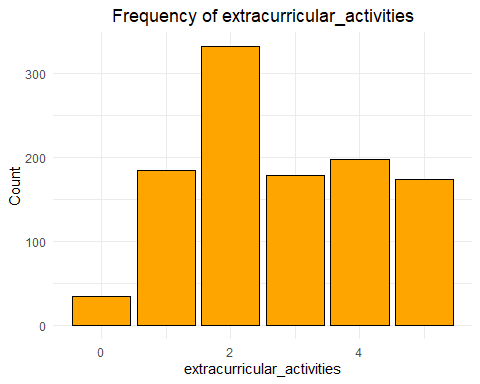
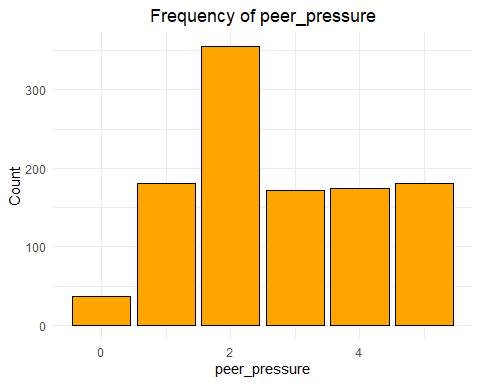
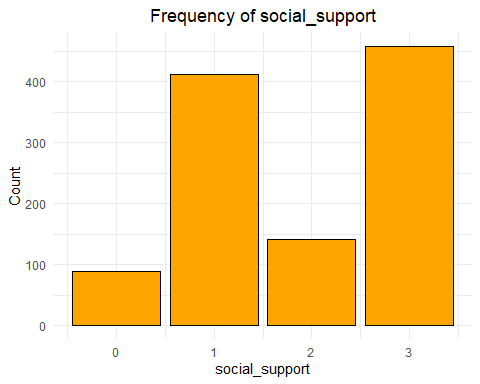
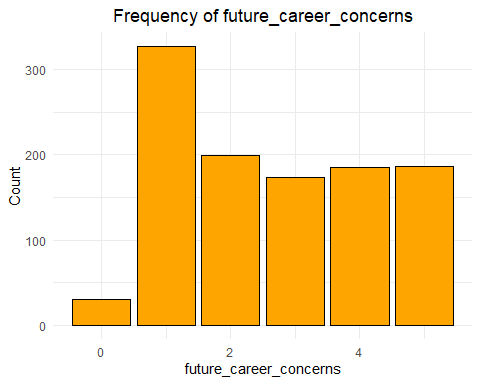
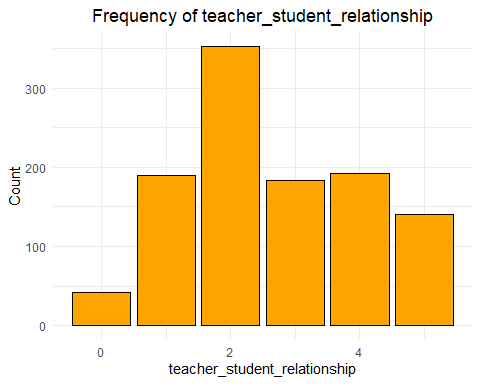
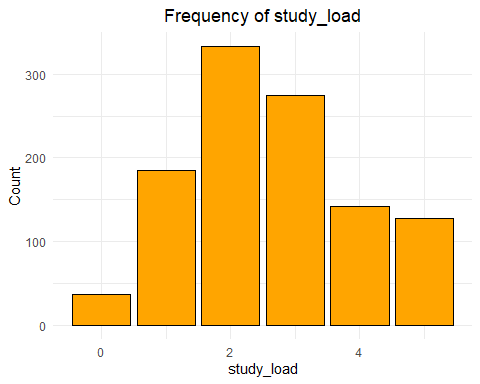
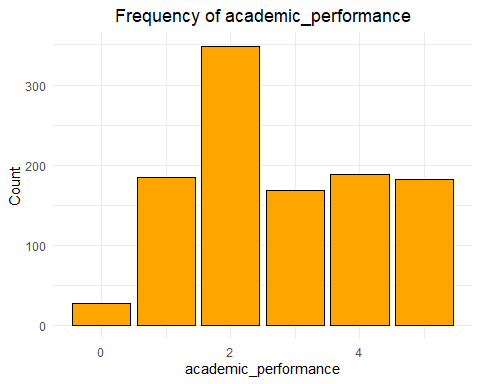
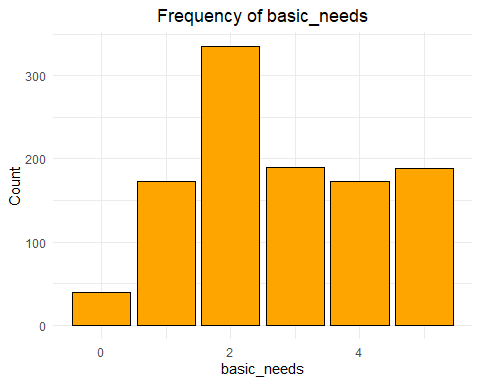
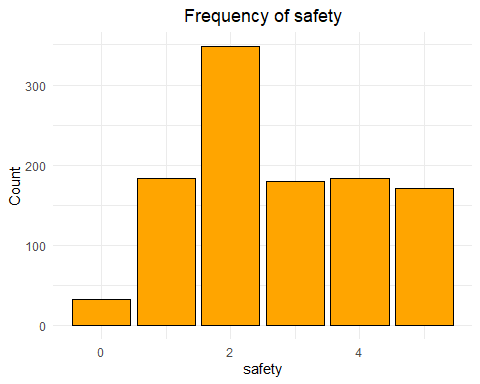
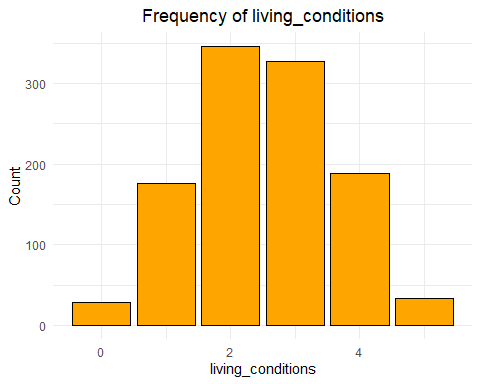
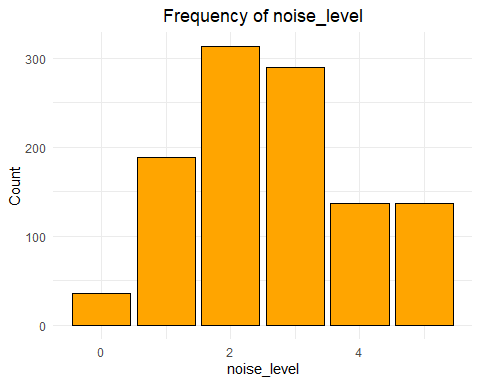
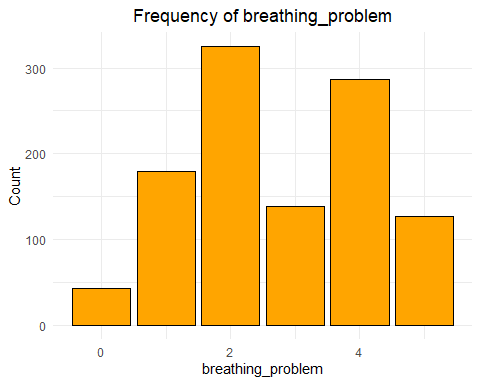
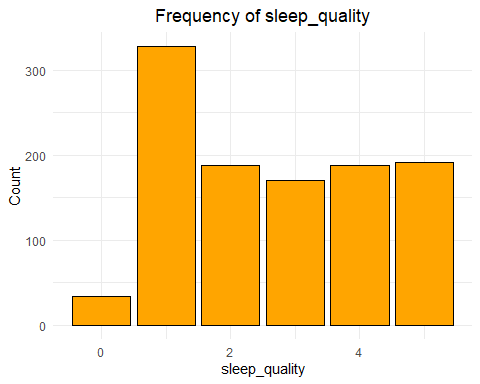
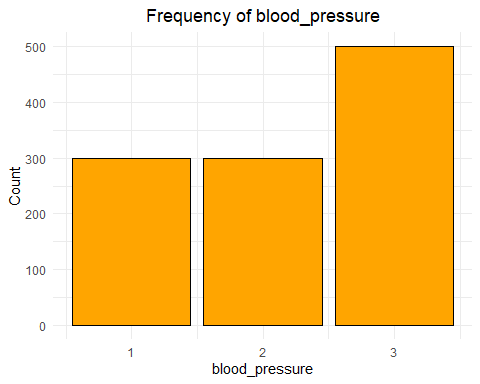
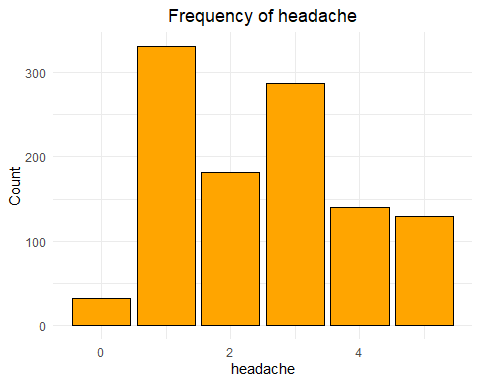
## *Creating and Interpreting Scatter Plots of Variables:*

library(ggplot2)  
  
continuous\_vars <- c("anxiety\_level", "self\_esteem", "depression")  
for (var in continuous\_vars) {  
 p <- ggplot(Stress\_level\_data, aes\_string(x = var)) +  
 geom\_histogram(binwidth = 2, color = "black", fill = "skyblue") +  
 labs(title = paste("Distribution of", var), x = var, y = "Frequency") +  
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5))  
   
 print(p)   
}

## Warning: `aes\_string()` was deprecated in ggplot2 3.0.0.  
## ℹ Please use tidy evaluation idioms with `aes()`.  
## ℹ See also `vignette("ggplot2-in-packages")` for more information.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



categorical\_vars <- c("headache", "blood\_pressure", "sleep\_quality", "breathing\_problem",   
 "noise\_level", "living\_conditions", "safety", "basic\_needs",   
 "academic\_performance", "study\_load", "teacher\_student\_relationship",   
 "future\_career\_concerns", "social\_support", "peer\_pressure",   
 "extracurricular\_activities", "bullying", "mental\_health\_history", "stress\_level")  
  
for (var in categorical\_vars) {  
 p <- ggplot(Stress\_level\_data, aes\_string(x = var)) +  
 geom\_bar(fill = "orange", color = "black") +  
 labs(title = paste("Frequency of", var), x = var, y = "Count") +  
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5))  
   
 print(p)   
}



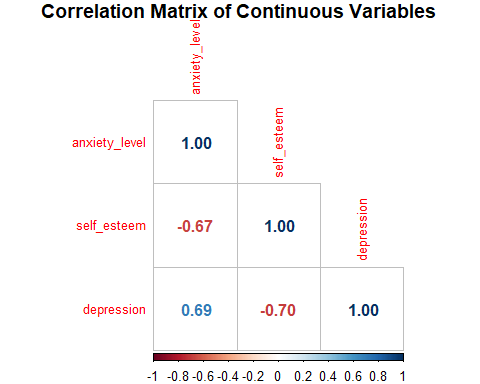
1. The students’ anxiety level is concentrated at a moderate level. Low and high values are rare.
2. The distribution graph of the self-esteem level is skewed to the left. Most of the students have a self-esteem level between 20-30.
3. Most of the students have a low-moderate depression level.
4. Most of the students have a low-moderate headache.
5. More than half of the students have a low-moderate blood pressure level.
6. Half of the students have a sleep level below the moderate level, while the remaining half have a sleep level above the moderate level.
7. Some of the students have moderate respiratory problems, while some have serious respiratory problems.
8. Most of the students are exposed to moderate noise.
9. Most of the students have moderate living conditions.
10. Most of the students have a high perception of security.
11. Most of the students basic needs are met at a moderate level.
12. Most of the students have medium level and above academic success.
13. Most of the students have medium level workload.
14. Most of the students have medium level teacher-student relationships.
15. Most of the students have low-moderate career anxiety.
16. About half of the students receive high social support.
17. Most of the students experience medium level peer pressure.
18. Most of the students participate in extracurricular activities.
19. Most of the students experience low-moderate level bullying.
20. Half of the students have a mental health history, but half do not have a mental health history.
21. Stress levels are equally distributed among the students as low, medium and high.

## *Examining Correlation Matrices:*

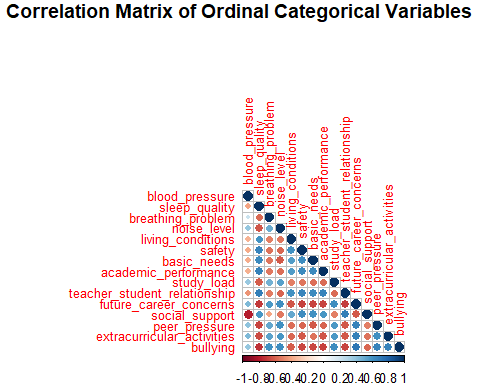
library(corrplot)

## corrplot 0.95 loaded

cor\_matrix <- cor(Stress\_level\_data[, c("anxiety\_level", "self\_esteem", "depression")],   
 use = "complete.obs")  
corrplot(cor\_matrix, method = "number", type = "lower",   
 tl.cex = 0.8, title = "Correlation Matrix of Continuous Variables",   
 mar = c(0, 0, 1, 0))



ordinal\_vars <- c("blood\_pressure", "sleep\_quality", "breathing\_problem",   
 "noise\_level", "living\_conditions", "safety",   
 "basic\_needs", "academic\_performance",   
 "study\_load", "teacher\_student\_relationship",   
 "future\_career\_concerns", "social\_support",   
 "peer\_pressure", "extracurricular\_activities", "bullying")  
ordinal\_cor\_matrix <- cor(Stress\_level\_data[, ordinal\_vars], method = "spearman", use = "complete.obs")  
corrplot(ordinal\_cor\_matrix, method = "circle", type = "lower",  
 tl.cex = 0.8, title = "Correlation Matrix of Ordinal Categorical Variables",   
 mar = c(0, 0, 1, 0))



The correlation coefficient takes value between -1 and +1. If the coefficient is close to +1, there is a strong positive relationship between the variables. If the coefficient is close to -1, there is a strong negative relationship between the variables. If the correlation coefficient is 0, there is no relationship between the two variables.

Since the relationship between Anxiety\_level and Depression is +0.69, as anxiety levels increase, depression levels increase.

Since the relationship between Self\_esteem and Depression is -0.70, as self-esteem levels increase, depression levels decrease.

Since the relationship between Anxiety\_level and Self\_esteem is -0.67, as self-esteem levels increase, anxiety levels decrease.

There is a significant positive correlation between the variables Study\_load and Academic\_performance. In other words, as study load increases, academic performance will also increase.

There is a positive relationship between the variables Social\_support and Future\_career\_concerns. As social support increases, students’ career concerns will also increase.

There is a positive relationship between the variables Teacher\_student\_relationship and Academic\_performance. This means that when the student’s relationship with the teacher increases, their academic success will also increase.

There is a strong negative relationship between the variables Bullying and Social\_support. If social support increases, bullying will decrease.

There is a negative relationship between Bullying and Academic\_performance. If bullying increases, academic success will decrease.

There is a negative relationship between Noise\_level and Sleep\_quality. If the noise level increases, sleep quality will decrease.

There is a positive relationship between the variables Study\_load and Future\_career\_concerns. If career concerns increase, study load will also increase.

## 3-Formal Modelling

Stress\_level\_data$stress\_level <- as.factor(Stress\_level\_data$stress\_level)  
head(Stress\_level\_data)

## # A tibble: 6 × 21  
## anxiety\_level self\_esteem mental\_health\_history depression headache  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 14 20 0 11 2  
## 2 15 8 1 15 5  
## 3 12 18 1 14 2  
## 4 16 12 1 15 4  
## 5 16 28 0 7 2  
## 6 20 13 1 21 3  
## # ℹ 16 more variables: blood\_pressure <dbl>, sleep\_quality <dbl>,  
## # breathing\_problem <dbl>, noise\_level <dbl>, living\_conditions <dbl>,  
## # safety <dbl>, basic\_needs <dbl>, academic\_performance <dbl>,  
## # study\_load <dbl>, teacher\_student\_relationship <dbl>,  
## # future\_career\_concerns <dbl>, social\_support <dbl>, peer\_pressure <dbl>,  
## # extracurricular\_activities <dbl>, bullying <dbl>, stress\_level <fct>

library(caret)

## Zorunlu paket yükleniyor: lattice

set.seed(42)  
trainIndex <- createDataPartition(Stress\_level\_data$stress\_level, p = 0.8, list = FALSE, times = 1)  
trainData <- Stress\_level\_data[trainIndex, ]  
testData <- Stress\_level\_data[-trainIndex, ]  
dim(trainData)

## [1] 882 21

dim(testData)

## [1] 218 21

In this code block, the stress\_level variable was converted to a categorical data type in order to establish the model, and 80% of the data set was separated as training and 20% as testing. In order to obtain the same results when the model was run repeatedly, a fixed seed value was assigned with set.seed(42). Then, the sizes of the training and testing data sets were checked.

library(rpart)

## Warning: package 'rpart' was built under R version 4.4.2

tree\_model <- rpart(stress\_level ~ ., data = trainData, method = "class")  
summary(tree\_model)

## Call:  
## rpart(formula = stress\_level ~ ., data = trainData, method = "class")  
## n= 882   
##   
## CP nsplit rel error xerror xstd  
## 1 0.4202401 0 1.0000000 1.0566038 0.02337913  
## 2 0.0100000 2 0.1595197 0.1715266 0.01615104  
##   
## Variable importance  
## blood\_pressure self\_esteem   
## 32 16   
## social\_support safety   
## 14 13   
## teacher\_student\_relationship basic\_needs   
## 13 13   
##   
## Node number 1: 882 observations, complexity param=0.4202401  
## predicted class=0 expected loss=0.6609977 P(node) =1  
## class counts: 299 287 296  
## probabilities: 0.339 0.325 0.336   
## left son=2 (641 obs) right son=3 (241 obs)  
## Primary splits:  
## blood\_pressure < 1.5 to the right, improve=226.3702, (0 missing)  
## teacher\_student\_relationship < 3.5 to the right, improve=208.8120, (0 missing)  
## basic\_needs < 3.5 to the right, improve=207.0417, (0 missing)  
## safety < 3.5 to the right, improve=206.4955, (0 missing)  
## academic\_performance < 3.5 to the right, improve=205.8858, (0 missing)  
##   
## Node number 2: 641 observations, complexity param=0.4202401  
## predicted class=0 expected loss=0.5335413 P(node) =0.7267574  
## class counts: 299 46 296  
## probabilities: 0.466 0.072 0.462   
## left son=4 (277 obs) right son=5 (364 obs)  
## Primary splits:  
## self\_esteem < 24.5 to the right, improve=200.9158, (0 missing)  
## blood\_pressure < 2.5 to the left, improve=200.8224, (0 missing)  
## social\_support < 2 to the right, improve=200.8224, (0 missing)  
## academic\_performance < 3.5 to the right, improve=193.7316, (0 missing)  
## basic\_needs < 3.5 to the right, improve=192.4618, (0 missing)  
## Surrogate splits:  
## blood\_pressure < 2.5 to the left, agree=0.952, adj=0.888, (0 split)  
## social\_support < 2 to the right, agree=0.952, adj=0.888, (0 split)  
## safety < 3.5 to the right, agree=0.927, adj=0.830, (0 split)  
## teacher\_student\_relationship < 3.5 to the right, agree=0.917, adj=0.809, (0 split)  
## basic\_needs < 3.5 to the right, agree=0.911, adj=0.794, (0 split)  
##   
## Node number 3: 241 observations  
## predicted class=1 expected loss=0 P(node) =0.2732426  
## class counts: 0 241 0  
## probabilities: 0.000 1.000 0.000   
##   
## Node number 4: 277 observations  
## predicted class=0 expected loss=0.06137184 P(node) =0.314059  
## class counts: 260 9 8  
## probabilities: 0.939 0.032 0.029   
##   
## Node number 5: 364 observations  
## predicted class=2 expected loss=0.2087912 P(node) =0.4126984  
## class counts: 39 37 288  
## probabilities: 0.107 0.102 0.791

In this code block, a decision tree model was created. All other independent variables were used to estimate the stress\_level variable. The train data set was used as the data set and the method = “class” parameter was determined since the classification process would be performed. Finally, with the summary(tree\_model) function, summary information of the decision tree created, information about the structure of the model and the variables used were obtained.

The model works on 882 observations in the train set and shows that personal health factors such as blood pressure and self-esteem are directly related to stress. High blood pressure and low self-esteem can increase stress levels. In addition, environmental factors such as social support and security also affect stress. Support from family, friends and society play an important role in coping with stress. The teacher-student relationship in the educational environment and meeting basic needs also affect stress levels. When the student-teacher relationship is good, stress levels decrease, and when it is bad, stress levels increase. When basic needs are met, stress levels decrease, and when they are not met, stress levels increase.

y\_pred <- predict(tree\_model, testData, type = "class")  
conf\_matrix <- confusionMatrix(y\_pred, testData$stress\_level)  
conf\_matrix

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1 2  
## 0 59 2 1  
## 1 0 59 0  
## 2 15 10 72  
##   
## Overall Statistics  
##   
## Accuracy : 0.8716   
## 95% CI : (0.8197, 0.9129)  
## No Information Rate : 0.3394   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.8072   
##   
## Mcnemar's Test P-Value : 2.215e-05   
##   
## Statistics by Class:  
##   
## Class: 0 Class: 1 Class: 2  
## Sensitivity 0.7973 0.8310 0.9863  
## Specificity 0.9792 1.0000 0.8276  
## Pos Pred Value 0.9516 1.0000 0.7423  
## Neg Pred Value 0.9038 0.9245 0.9917  
## Prevalence 0.3394 0.3257 0.3349  
## Detection Rate 0.2706 0.2706 0.3303  
## Detection Prevalence 0.2844 0.2706 0.4450  
## Balanced Accuracy 0.8882 0.9155 0.9069

In this code block, tree\_model was used to make predictions on the test data set and the results were stored in the y\_pred variable. Using the confusionMatrix function, the predicted values were compared with the actual values. As a result of these operations, a confusion matrix was created to evaluate the performance of the model.

The model correctly classified 87.16% of all observations. When looking at the p-value, it is seen that the accuracy of the model is much better than a random result. The Kappa value shows the classification success. A Kappa value above 0.80 shows that the model provides a strong fit and a reliable model is preferred. Sensitivity: 79.73% When looking at the sensitivity value for Model 0, it is seen that 79.73% of the real Class 0 observations are correctly classified. Sensitivity: 83.10% When looking at the sensitivity value for Model 1, it is seen that 83.10% of the real Class 1 observations are correctly classified. Sensitivity: 98.63% When looking at the sensitivity value for Model 2, it is seen that 98.63% of the real Class 2 observations are correctly classified.

rpart.plot(tree\_model,   
 type = 5,   
 extra = 104,   
 main = "Decision Tree Visualization",  
 box.palette = list("lightblue", "lightgreen", "lightpink"),   
 shadow.col = "gray",   
 branch.lty = 3,   
 nn = TRUE,   
 fallen.leaves = TRUE,   
 digits = 3)

metin, ekran görüntüsü, diyagram, tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldu

In this code block, the decision tree model named tree\_model is visualized using the rpart.plot function. Node type is determined with type = 5. Class and probability values are shown on the leaves with extra = 104. The graphic title is assigned as “Decision Tree Visualization”. Boxes are colored with box.palette. Shadow.col is determined as gray for shadows to make the visualization look better. The line type of the branches is determined as dashed line with branch.lty = 3. Node numbers are shown with nn = TRUE. Leaf nodes are aligned horizontally with fallen.leaves = TRUE and numerical values are shown with 3 digits with digits = 3.

# 4-Interpretation

In the decision tree, the variables determined by the model as the most effective were used. These are blood pressure and self-esteem variables. If the student's blood pressure is less than 2, there is a 100% chance that the student is at a moderate level of stress. If the student's blood pressure is 2 or higher and their self-esteem is 25 or higher, there is a 93.9% chance that they are at a low level of stress. There is a 3.2% chance that they are at a moderate level of stress and a 2.9% chance that they are at a high level of stress. If the student's blood pressure is 2 or higher and their self-esteem is less than 25, there is a 79.1% chance that they are at a high level of stress. There is a 10.2% chance that they are at a moderate level of stress and a 10.7% chance that they are at a low level of stress.