1.1 Data set import and export

```
# Read the csv file.
data <- read.csv(file = "input.csv")

# Display contents.
data

# Write data to file.
write.csv(data, "output.csv")
```

1.2 Data exploration and visualization

```
data <- read.csv(file = "input.csv")

# Exploration
head(data)
names(data)
attributes(data)
summary(data)
mean(data$Sepal.Length)
median(data$Sepal.Width)
range(data$Petal.Length)
quantile(data$Sepal.Length)
var(data$Sepal.Length)
table(data$Species)

#Covarience and Correlation
cov(data$Sepal.Length, data$Petal.Length)
cor(data$Sepal.Length, data$Petal.Length)

#Visualization
hist(data$Sepal.Length)
```

```
plot(density(data$Sepal.Length))
pie(table(data$Species))
barplot(table(data$Species))
plot(data$Petal.Length)
```

1.3 Generate association rule using apriori and visualize them

```
library(arules)
library(arulesViz)

patterns = random.patterns(nItems = 1000)
trans = random.transactions(nItems = 1000, nTrans = 1000, method = "agrawal",
patterns = patterns)
image(trans)
rules = apriori(trans, parameter=list(support=0.01, confidence=0.5))
inspect(rules)
plot(rules, method="grouped")
```

1.4 Construct decision tree and naïve Bayesian classifiers. Visualize and compare the results for accuracy

```
install.packages("party")
install.packages("e1071")
install.packages("caret")

#load libraries
```

```
library(party)  #Contains the decision tree functions
library(caret)  #Contains confusion matrix functions
library(e1071)  #Contains the naive bayes functions

#data <- read.csv(file = "input.csv")
#View(data)

#We need a large dataset for modeling the decision tree so inform the teacher and
use a built in dataset
data <- readingSkills
#Split dataset into test and train
index <- sample(2, nrow(data), replace=TRUE, prob=c(0.7,0.3))
train <- data[index==1,]
test <- data[index==2,]

#Select the dependent and independent features
features <- nativeSpeaker ~ age + shoeSize + score

#Obtain a decision tree model
model <- ctree(features, data=train)

#Plot the model
plot(model)

#Evaluate model on test data
test_predictions <- predict(model, newdata=test)
confusionMatrix(test_predictions, test$nativeSpeaker, positive="yes")

#Obtain a decision tree model
model2 <- naiveBayes(features, data=train)

#Model Summary
print(model2)

#Evaluate model on test data
test_predictions2 <- predict(model2, newdata = test)
confusionMatrix(test_predictions2, test$nativeSpeaker, positive="yes")
```

1.5 Perform linear regression on a dataset and visualize the results.

```
#Input dataset
data <- read.csv("input.csv")

#Selecting columns
age <- data[1:7,1:1]
circum <- data[1:7,2:2]

plot(age, circum, xlab = "age", ylab = "circumference")

data <- data.frame(age, circum)
model = lm(circum ~ age, data)
summary(model)

abline(model, cex=1)

new_data <- data.frame(age = 700)
result <- predict(model, new_data)
print(result)
```

1.6 Build clusters using K-means and Hierarchical clustering and visualize the results

```
#load libraries
library(cluster)
library(factoextra)


#Input dataset
```

```
data <- read.csv(“input.csv")

#Making species column NULL
data$Species<-NULL

d<- scale(dist(data,method = "euclidian"))
#KMeans
kfit <- kmeans(d,3)
#HClustering
hfit <- hkmeans(d, 3)
fviz_cluster(kfit, data)
fviz_cluster(hfit, data)
```