Q2

Gowtham P

2. Using the "Marketing Customer Value Analysis" dataset, complete the following tasks with proper analysis and interpretation:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
#### 1. Load Data ####
df <- read.csv("D:/Data Science for Marketing-I& 2/dataset/WA_Fn-UseC_-Marketing-Customer-Value-
Analysis.csv")
```

i. Create a new column named "Engaged" by transforming the categorical values in the "Response" variable into numerical values. Why is this transformation important?

```
## 2.1. Response Variable: Response
df$Engaged <- rep(0,nrow(df))
df$Engaged[df$Response=='Yes']=1
```

Interpertation:

This transformation converts categorical responses into numeric values.

ii. Explain the process of creating dummy variables from the categorical features ['Sales Channel', 'Vehicle Size', 'Vehicle Class', 'Policy', 'Policy Type', 'Employment Status', 'Marital Status', 'Education', 'Coverage'] in the dataset. What function is used, and what is its significance in preparing the data for modeling?

```
## 2.2. Categorical Features
categoricalVars = c(
  'Sales.Channel', 'Vehicle.Size', 'Vehicle.Class', 'Policy', 'Policy.Type',
  'EmploymentStatus', 'Marital.Status', 'Education', 'Coverage', 'Gender'
)

encodedDF <- model.matrix(~.-1, df[categoricalVars])
```

Interpertation:

Dummy encoding converts categorical features into numerical format, allowing models to process them effectively.

iii. Combine all features into a single dataset for model building.

```
## 2.3. Continuous Features
continuousFeatures <- c(
  'Customer.Lifetime.Value', 'Income', 'Monthly.Premium.Auto',
  'Months.Since.Last.Claim', 'Months.Since.Policy.Inception',
  'Number.of.Open.Complaints', 'Number.of.Policies', 'Total.Claim.Amount'
)

encodedDF <- cbind(encodedDF, df[continuousFeatures])
```

Interpertation:

This ensures all relevant features, including numerical and dummy variables, are included for modeling.

iv. Split the dataset into a 70:30 ratio for training and test data, and build a Random Forest model using the training data.

```
# install.packages('caTools')
library(caTools)
```

```
set.seed(123)
sample <- sample.split(df$Customer, SplitRatio = .7)
```

```
trainX <- as.matrix(subset(encodedDF, sample == TRUE))
trainY <- as.double(as.matrix(subset(df$Engaged, sample == TRUE)))
```

```
testX <- as.matrix(subset(encodedDF, sample == FALSE))
testY <- as.double(as.matrix(subset(df$Engaged, sample == FALSE)))
```

Interpertation:

The dataset is split into training (70%) and test (30%) sets, and a Random Forest model is trained to predict engagement.

v. After fitting the Random Forest model, how can you interpret the output of rf_model.feature_importances_? What does this tell you about the features in the dataset?

- Training

```r
# install.packages('randomForest')
library(randomForest)
```

```
## randomForest 4.7-1.2
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
## 
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
## 
##     margin
```

```
## The following object is masked from 'package:dplyr':
## 
##     combine
```

```r
rfModel <- randomForest(x=trainX, y=factor(trainY), ntree=200, maxnodes=24)
```

```r
# - Individual Tree Predictions
getTree(rfModel, 1)
```

```
##    left daughter right daughter split var split point status prediction
## 1             2              3         3      0.5000      1          0
## 2             4              5        36  10592.0000      1          0
## 3             6              7        37     95.5000      1          0
## 4             8              9        38     34.5000      1          0
## 5            10             11        22      0.5000      1          0
## 6            12             13        29      0.5000      1          0
## 7            14             15        34      0.5000      1          0
## 8            16             17         1      0.5000      1          0
## 9            18             19        40      2.5000      1          0
## 10           20             21        38     22.5000      1          0
## 11           22             23        37    275.0000      1          0
## 12           24             25        26      0.5000      1          0
## 13            0              0         0      0.0000     -1          1
## 14           26             27        42    555.2057      1          0
## 15           28             29        10      0.5000      1          0
## 16           30             31        35   2019.6719      1          0
## 17           32             33        29      0.5000      1          0
## 18           34             35        35  21033.9521      1          0
## 19            0              0         0      0.0000     -1          2
## 20           36             37        42    254.2915      1          0
## 21           38             39        42    628.5443      1          0
## 22           40             41        40      1.5000      1          0
## 23           42             43        38     30.5000      1          0
## 24           44             45        38     11.5000      1          0
## 25           46             47        31      0.5000      1          0
## 26            0              0         0      0.0000     -1          1
## 27            0              0         0      0.0000     -1          1
## 28            0              0         0      0.0000     -1          1
## 29            0              0         0      0.0000     -1          1
## 30            0              0         0      0.0000     -1          2
## 31            0              0         0      0.0000     -1          1
## 32            0              0         0      0.0000     -1          1
## 33            0              0         0      0.0000     -1          1
## 34            0              0         0      0.0000     -1          1
## 35            0              0         0      0.0000     -1          2
## 36            0              0         0      0.0000     -1          1
## 37            0              0         0      0.0000     -1          1
## 38            0              0         0      0.0000     -1          1
## 39            0              0         0      0.0000     -1          1
## 40            0              0         0      0.0000     -1          1
## 41            0              0         0      0.0000     -1          1
## 42            0              0         0      0.0000     -1          1
## 43            0              0         0      0.0000     -1          2
## 44            0              0         0      0.0000     -1          1
## 45            0              0         0      0.0000     -1          1
## 46            0              0         0      0.0000     -1          1
## 47            0              0         0      0.0000     -1          1
```

```r
predict<-predict(rfModel, trainX, predict.all=TRUE)$individual
head(predict, 5)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14]
## 1   "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"   "0"   "0"   "0"   "0"
## 3   "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"   "0"   "0"   "0"   "0"
## 6   "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"   "0"   "0"   "0"   "0"
## 7   "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"  "0"   "0"   "0"   "0"   "0"
## 9   "0"  "0"  "0"  "0"  "0"  "1"  "0"  "0"  "1"  "0"   "0"   "0"   "0"   "0"
##      [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25] [,26] [,27]
## 1   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 3   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 6   "0"   "1"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 7   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 9   "0"   "0"   "0"   "0"   "1"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
##      [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37] [,38] [,39] [,40]
## 1   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 3   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 6   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 7   "0"   "1"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 9   "0"   "0"   "0"   "0"   "1"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
##      [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48] [,49] [,50] [,51] [,52] [,53]
## 1   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 3   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 6   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 7   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 9   "1"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
##      [,54] [,55] [,56] [,57] [,58] [,59] [,60] [,61] [,62] [,63] [,64] [,65] [,66]
## 1   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 3   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 6   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 7   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 9   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "1"   "1"   "0"   "0"   "0"
##      [,67] [,68] [,69] [,70] [,71] [,72] [,73] [,74] [,75] [,76] [,77] [,78] [,79]
## 1   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 3   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 6   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 7   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 9   "0"   "0"   "0"   "0"   "0"   "1"   "0"   "1"   "0"   "0"   "0"   "0"   "0"
##      [,80] [,81] [,82] [,83] [,84] [,85] [,86] [,87] [,88] [,89] [,90] [,91] [,92]
## 1   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 3   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 6   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
## 7   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "1"   "0"
## 9   "1"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"
##      [,93] [,94] [,95] [,96] [,97] [,98] [,99] [,100] [,101] [,102] [,103] [,104]
## 1   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"    "0"    "0"    "0"    "0"
## 3   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"    "0"    "0"    "0"    "0"
## 6   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"    "0"    "0"    "0"    "0"
## 7   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "0"    "0"    "0"    "0"    "0"
## 9   "0"   "1"   "0"   "0"   "0"   "0"   "0"   "0"    "0"    "0"    "0"    "0"
##      [,105] [,106] [,107] [,108] [,109] [,110] [,111] [,112] [,113] [,114] [,115]
## 1   "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 3   "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 6   "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
```

```
## 7 "1"   "0"   "0"   "0"   "0"   "0"   "0"   "0"   "1"   "0"   "0"
## 9 "0"   "0"   "0"   "0"   "0"   "1"   "0"   "0"   "0"   "0"   "0"
##   [,116] [,117] [,118] [,119] [,120] [,121] [,122] [,123] [,124] [,125] [,126]
## 1 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 3 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 6 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 7 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "1"    "0"
## 9 "0"    "0"    "1"    "1"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
##   [,127] [,128] [,129] [,130] [,131] [,132] [,133] [,134] [,135] [,136] [,137]
## 1 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 3 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 6 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 7 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 9 "0"    "0"    "0"    "0"    "1"    "0"    "0"    "1"    "0"    "0"    "0"
##   [,138] [,139] [,140] [,141] [,142] [,143] [,144] [,145] [,146] [,147] [,148]
## 1 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 3 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 6 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 7 "0"    "0"    "0"    "0"    "1"    "0"    "0"    "0"    "0"    "0"    "0"
## 9 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
##   [,149] [,150] [,151] [,152] [,153] [,154] [,155] [,156] [,157] [,158] [,159]
## 1 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 3 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 6 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 7 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 9 "0"    "0"    "1"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
##   [,160] [,161] [,162] [,163] [,164] [,165] [,166] [,167] [,168] [,169] [,170]
## 1 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 3 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 6 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 7 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 9 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "1"    "0"    "0"
##   [,171] [,172] [,173] [,174] [,175] [,176] [,177] [,178] [,179] [,180] [,181]
## 1 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 3 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 6 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 7 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 9 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
##   [,182] [,183] [,184] [,185] [,186] [,187] [,188] [,189] [,190] [,191] [,192]
## 1 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 3 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 6 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 7 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 9 "0"    "1"    "0"    "0"    "0"    "0"    "0"    "0"    "0"    "1"    "0"
##   [,193] [,194] [,195] [,196] [,197] [,198] [,199] [,200]
## 1 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 3 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 6 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 7 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
## 9 "0"    "0"    "0"    "0"    "0"    "0"    "0"    "0"
```

```
# - Feature Importances
importance(rfModel)
```

```
##                                 MeanDecreaseGini
## Sales.ChannelAgent                     8.02001748
## Sales.ChannelBranch                    1.05469894
## Sales.ChannelCall Center               2.50733881
## Sales.ChannelWeb                       0.66033914
## Vehicle.SizeMedsize                    1.33342875
## Vehicle.SizeSmall                      2.32321803
## Vehicle.ClassLuxury Car                0.12111307
## Vehicle.ClassLuxury SUV                0.16559466
## Vehicle.ClassSports Car                0.75678766
## Vehicle.ClassSUV                       1.44755721
## Vehicle.ClassTwo-Door Car              0.63205333
## PolicyCorporate L2                     0.20550321
## PolicyCorporate L3                     0.32576332
## PolicyPersonal L1                      0.34220645
## PolicyPersonal L2                      0.32137880
## PolicyPersonal L3                      0.31673264
## PolicySpecial L1                       0.03313553
## PolicySpecial L2                       0.19816534
## PolicySpecial L3                       0.13810589
## Policy.TypePersonal Auto               0.17401956
## Policy.TypeSpecial Auto                0.18957447
## EmploymentStatusEmployed               6.79196992
## EmploymentStatusMedical Leave          1.09876895
## EmploymentStatusRetired               70.14382562
## EmploymentStatusUnemployed             5.45579240
## Marital.StatusMarried                  3.49568328
## Marital.StatusSingle                   4.51379169
## EducationCollege                       0.68710303
## EducationDoctor                        0.74881960
## EducationHigh School or Below          0.83553220
## EducationMaster                        1.37124989
## CoverageExtended                       0.97348420
## CoveragePremium                        0.84141457
## GenderM                                0.84642049
## Customer.Lifetime.Value               11.20422547
## Income                                15.39870419
## Monthly.Premium.Auto                   9.20421521
## Months.Since.Last.Claim                5.00477312
## Months.Since.Policy.Inception          9.47997380
## Number.of.Open.Complaints              2.07664234
## Number.of.Policies                     3.27483213
## Total.Claim.Amount                    13.04214066
```

Interpertation:

Feature importance scores indicate which variables most influence engagement, helping in feature selection and understanding key drivers.

vi. Discuss how you would evaluate the performance of the Random Forest model after making predictions on the training and test set. Use all possible metrics, and explain why they are important for assessing model effectiveness.

```
inSamplePreds <- as.double(predict(rfModel, trainX)) - 1
outSamplePreds <- as.double(predict(rfModel, testX)) - 1
```

```
# - Accuracy, Precision, and Recall
inSampleAccuracy <- mean(trainY == inSamplePreds)
outSampleAccuracy <- mean(testY == outSamplePreds)
print(sprintf('In-Sample Accuracy: %0.4f', inSampleAccuracy))
```

```
## [1] "In-Sample Accuracy: 0.8713"
```

```
print(sprintf('Out-Sample Accuracy: %0.4f', outSampleAccuracy))
```

```
## [1] "Out-Sample Accuracy: 0.8701"
```

```
inSamplePrecision <- sum(inSamplePreds & trainY) / sum(inSamplePreds)
outSamplePrecision <- sum(outSamplePreds & testY) / sum(outSamplePreds)
print(sprintf('In-Sample Precision: %0.4f', inSamplePrecision))
```

```
## [1] "In-Sample Precision: 0.9898"
```

```
print(sprintf('Out-Sample Precision: %0.4f', outSamplePrecision))
```

```
## [1] "Out-Sample Precision: 0.9714"
```

```
inSampleRecall <- sum(inSamplePreds & trainY) / sum(trainY)
outSampleRecall <- sum(outSamplePreds & testY) / sum(testY)
print(sprintf('In-Sample Recall: %0.4f', inSampleRecall))
```

```
## [1] "In-Sample Recall: 0.1055"
```

```
print(sprintf('Out-Sample Recall: %0.4f', outSampleRecall))
```

```
## [1] "Out-Sample Recall: 0.0874"
```

- ROC & AUC

```
# install.packages('ROCR')
library(ROCR)
```
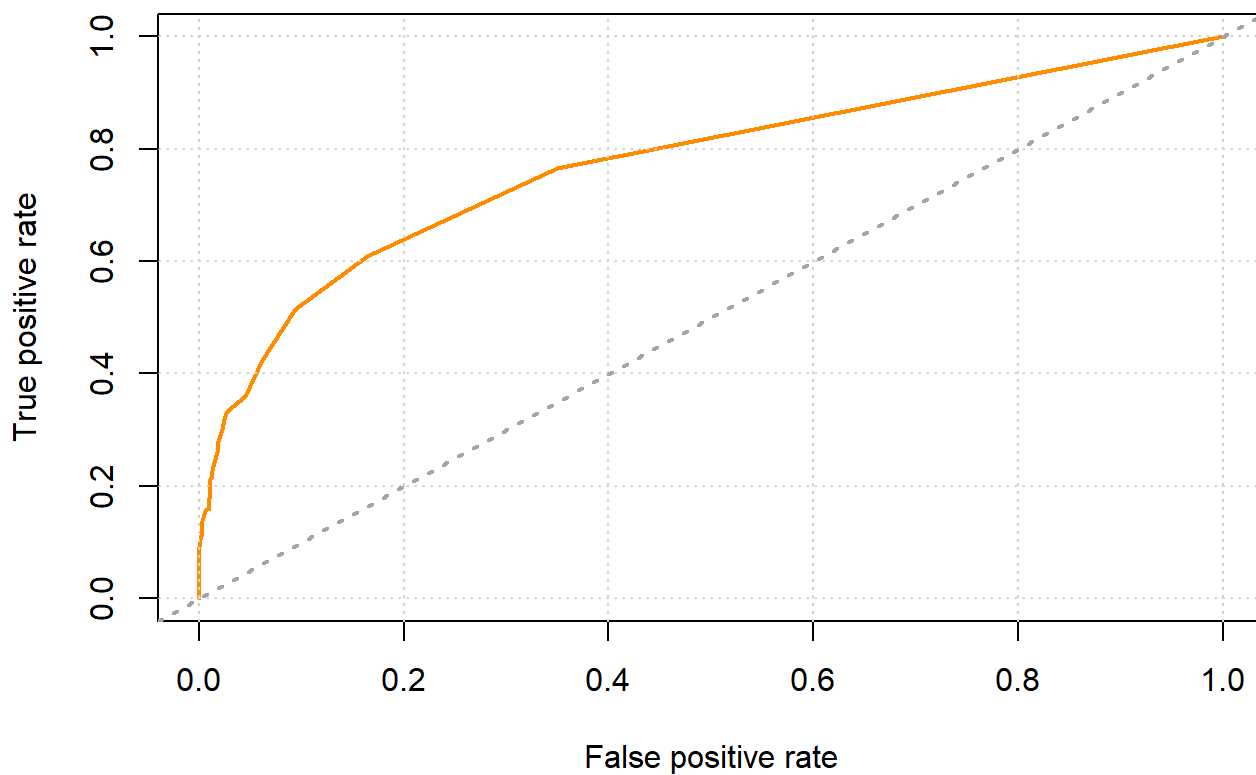
```
inSamplePredProbs <- as.double(predict(rfModel, trainX, type='prob')[,2])
outSamplePredProbs <- as.double(predict(rfModel, testX, type='prob')[,2])

pred <- prediction(outSamplePredProbs, testY)
perf <- performance(pred, measure = "tpr", x.measure = "fpr")
auc <- performance(pred, measure='auc')@y.values[[1]]

plot(
  perf,
  main = sprintf('Random Forest Model ROC Curve (AUC: %0.2f)', auc),
  col = 'darkorange',
  lwd = 2
)
grid()  # Call grid() separately
abline(a = 0, b = 1, col = 'darkgray', lty = 3, lwd = 2)
```

Random Forest Model ROC Curve (AUC: 0.78)

Interpertation:

Accuracy, confusion matrix, and AUC-ROC score assess how well the model predicts engagement, ensuring reliability in business decisions.