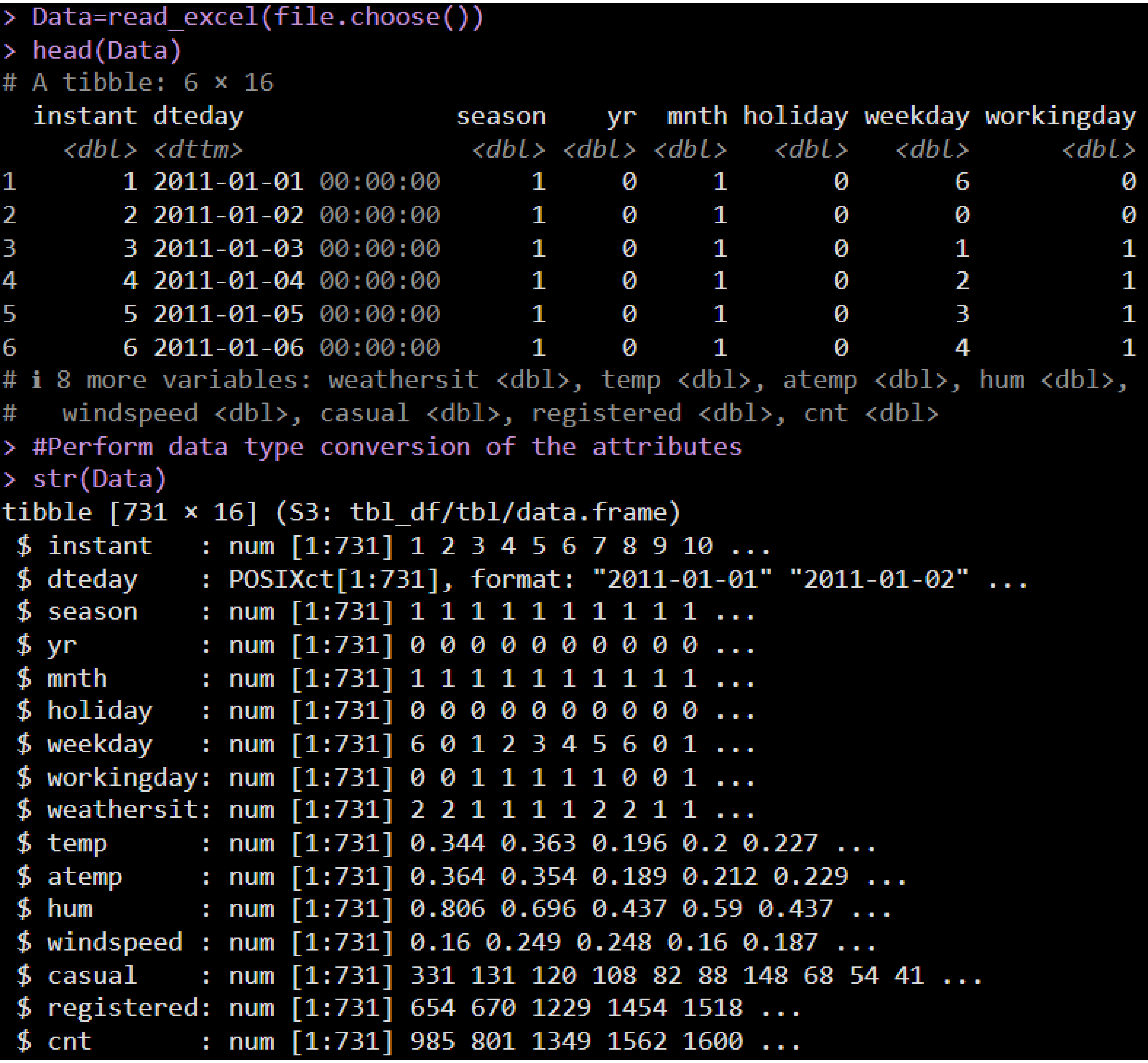
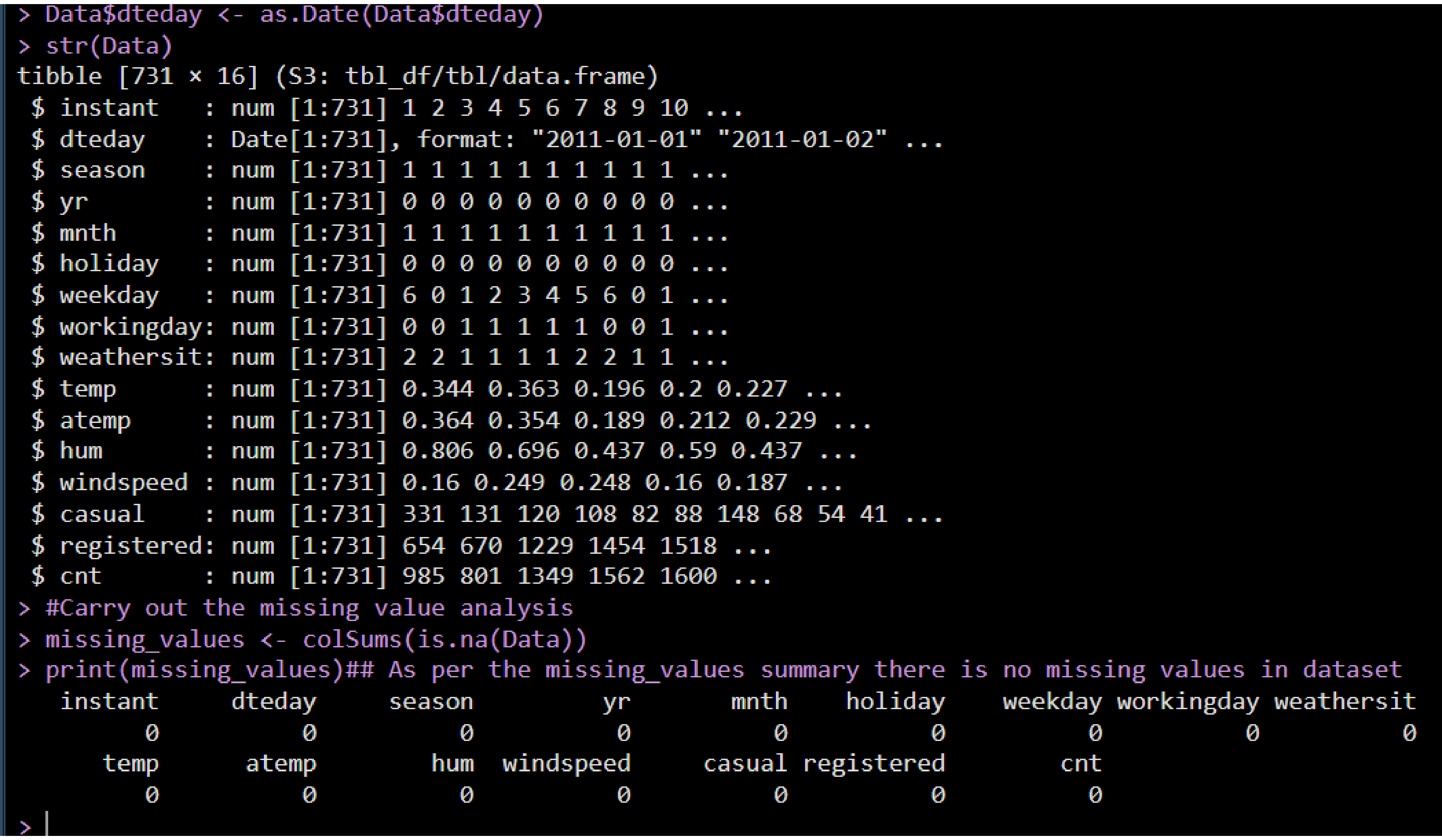
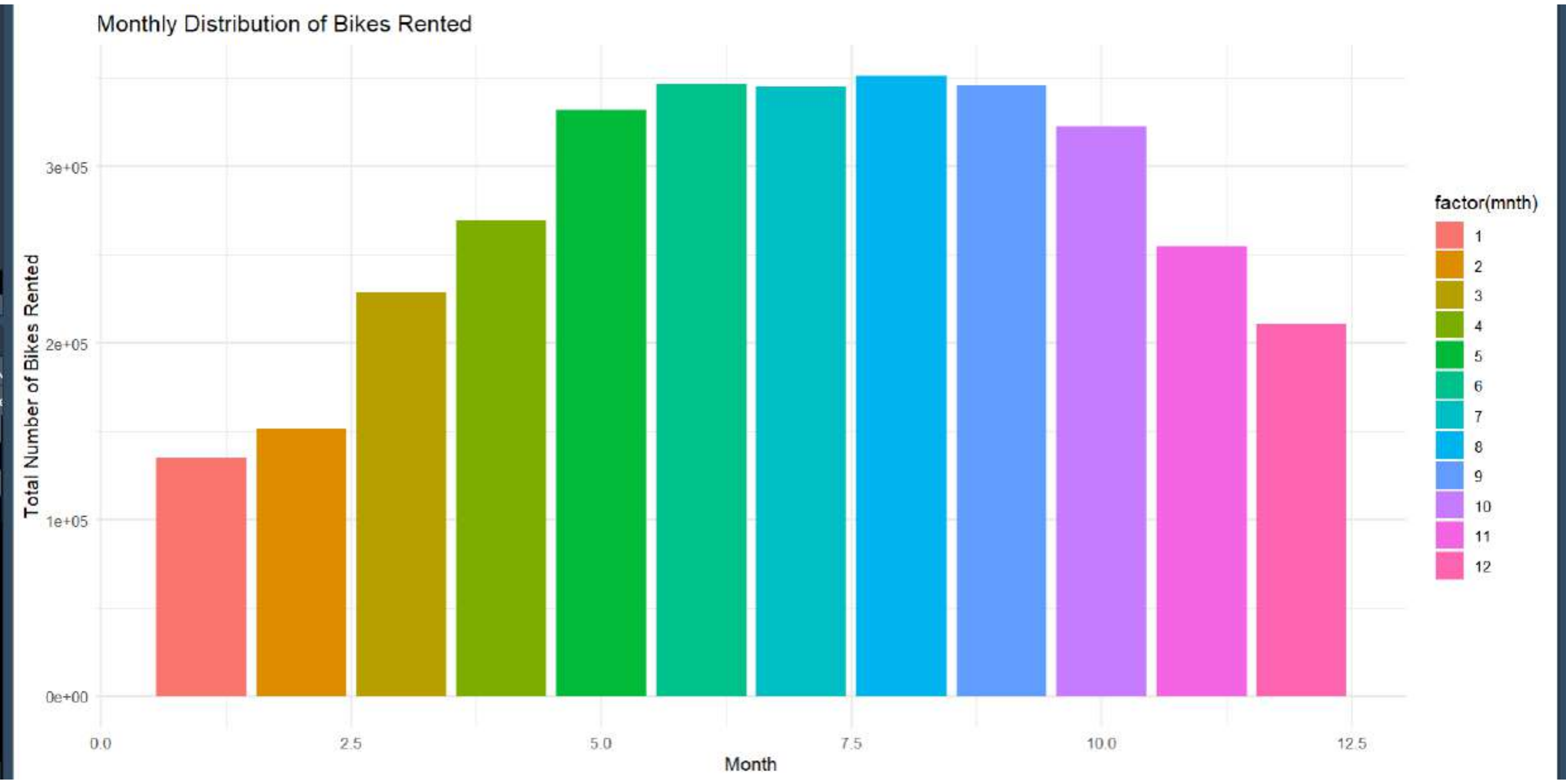
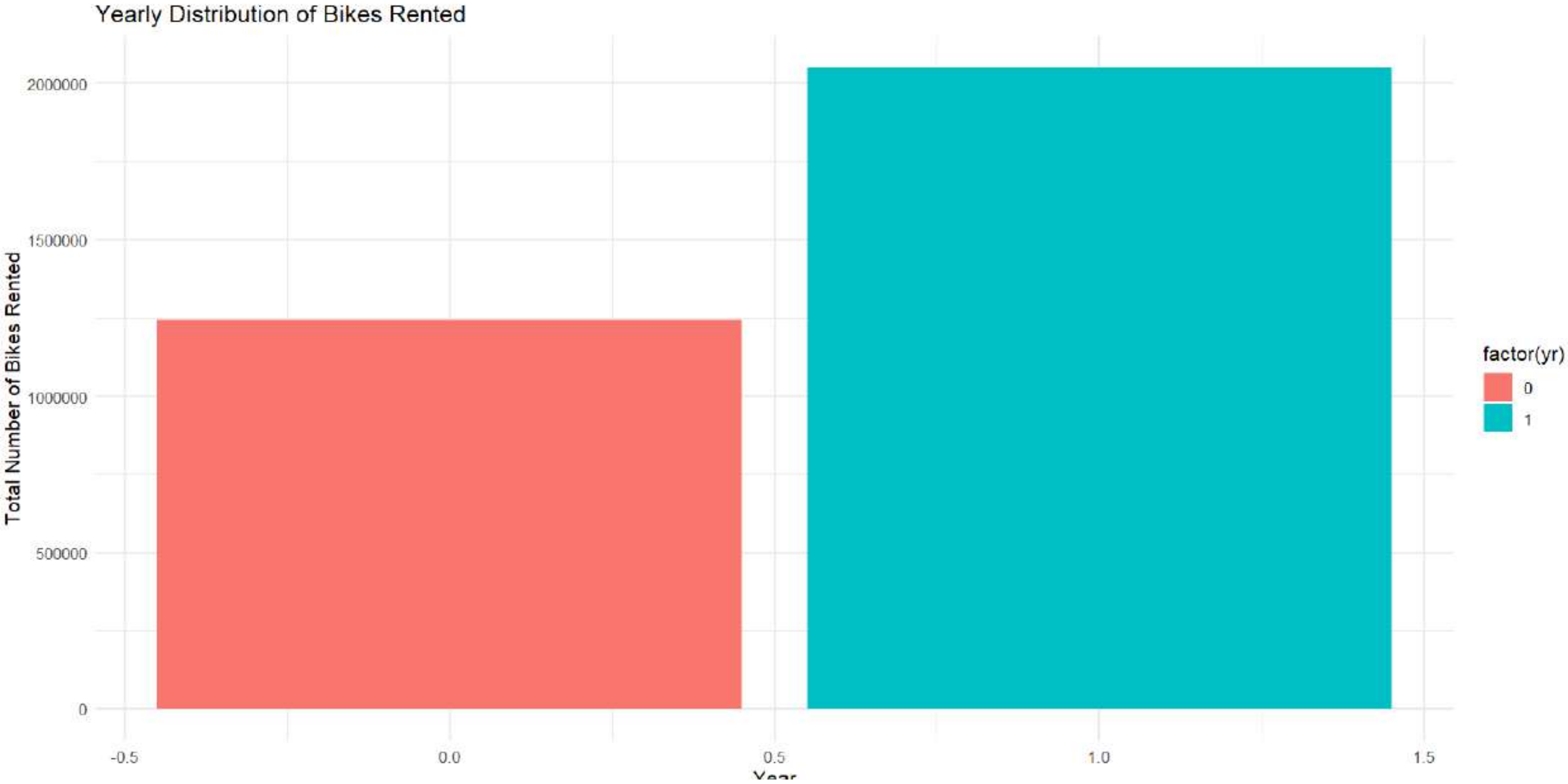


```
> Data=read_excel(file.choose())
> head(Data)
# A tibble: 6 × 16
  instant dteday              season    yr  mnth holiday weekday workingday
    <dbl> <dttm>               <dbl> <dbl> <dbl>   <dbl>   <dbl>      <dbl>
1       1 2011-01-01 00:00:00      1     0     1       0       6          0
2       2 2011-01-02 00:00:00      1     0     1       0       0          0
3       3 2011-01-03 00:00:00      1     0     1       0       1          1
4       4 2011-01-04 00:00:00      1     0     1       0       2          1
5       5 2011-01-05 00:00:00      1     0     1       0       3          1
6       6 2011-01-06 00:00:00      1     0     1       0       4          1
# ℹ 8 more variables: weathersit <dbl>, temp <dbl>, atemp <dbl>, hum <dbl>,
#   windspeed <dbl>, casual <dbl>, registered <dbl>, cnt <dbl>
> #Perform data type conversion of the attributes
> str(Data)
tibble [731 × 16] (S3: tbl_df/tbl/data.frame)
 $ instant   : num [1:731] 1 2 3 4 5 6 7 8 9 10 ...
 $ dteday    : POSIXct[1:731], format: "2011-01-01" "2011-01-02" ...
 $ season    : num [1:731] 1 1 1 1 1 1 1 1 1 1 ...
 $ yr        : num [1:731] 0 0 0 0 0 0 0 0 0 0 ...
 $ mnth      : num [1:731] 1 1 1 1 1 1 1 1 1 1 ...
 $ holiday   : num [1:731] 0 0 0 0 0 0 0 0 0 0 ...
 $ weekday   : num [1:731] 6 0 1 2 3 4 5 6 0 1 ...
 $ workingday: num [1:731] 0 0 1 1 1 1 1 0 0 1 ...
 $ weathersit: num [1:731] 2 2 1 1 1 1 2 2 1 1 ...
 $ temp      : num [1:731] 0.344 0.363 0.196 0.2 0.227 ...
 $ atemp     : num [1:731] 0.364 0.354 0.189 0.212 0.229 ...
 $ hum       : num [1:731] 0.806 0.696 0.437 0.59 0.437 ...
 $ windspeed : num [1:731] 0.16 0.249 0.248 0.16 0.187 ...
 $ casual    : num [1:731] 331 131 120 108 82 88 148 68 54 41 ...
 $ registered: num [1:731] 654 670 1229 1454 1518 ...
 $ cnt       : num [1:731] 985 801 1349 1562 1600 ...
```

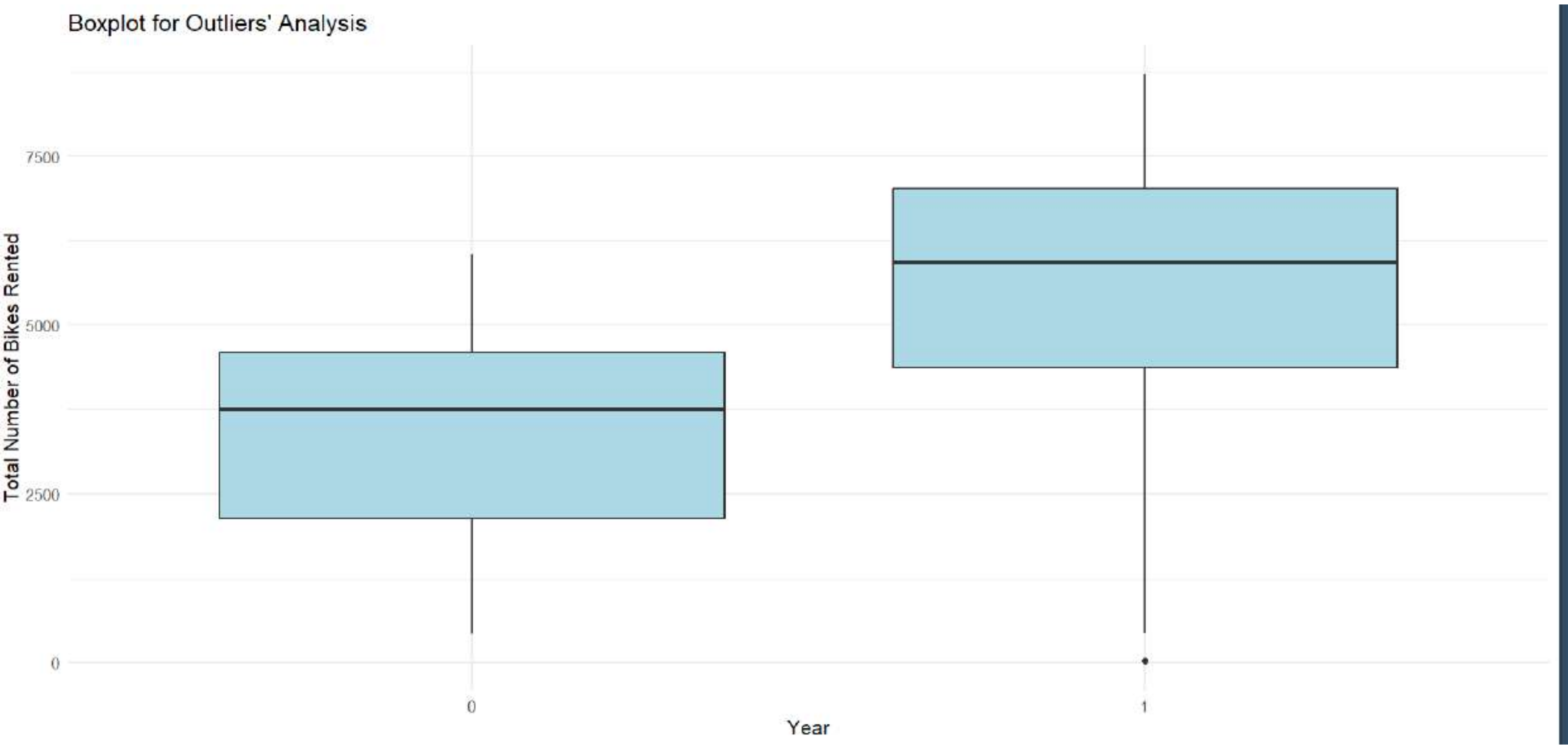


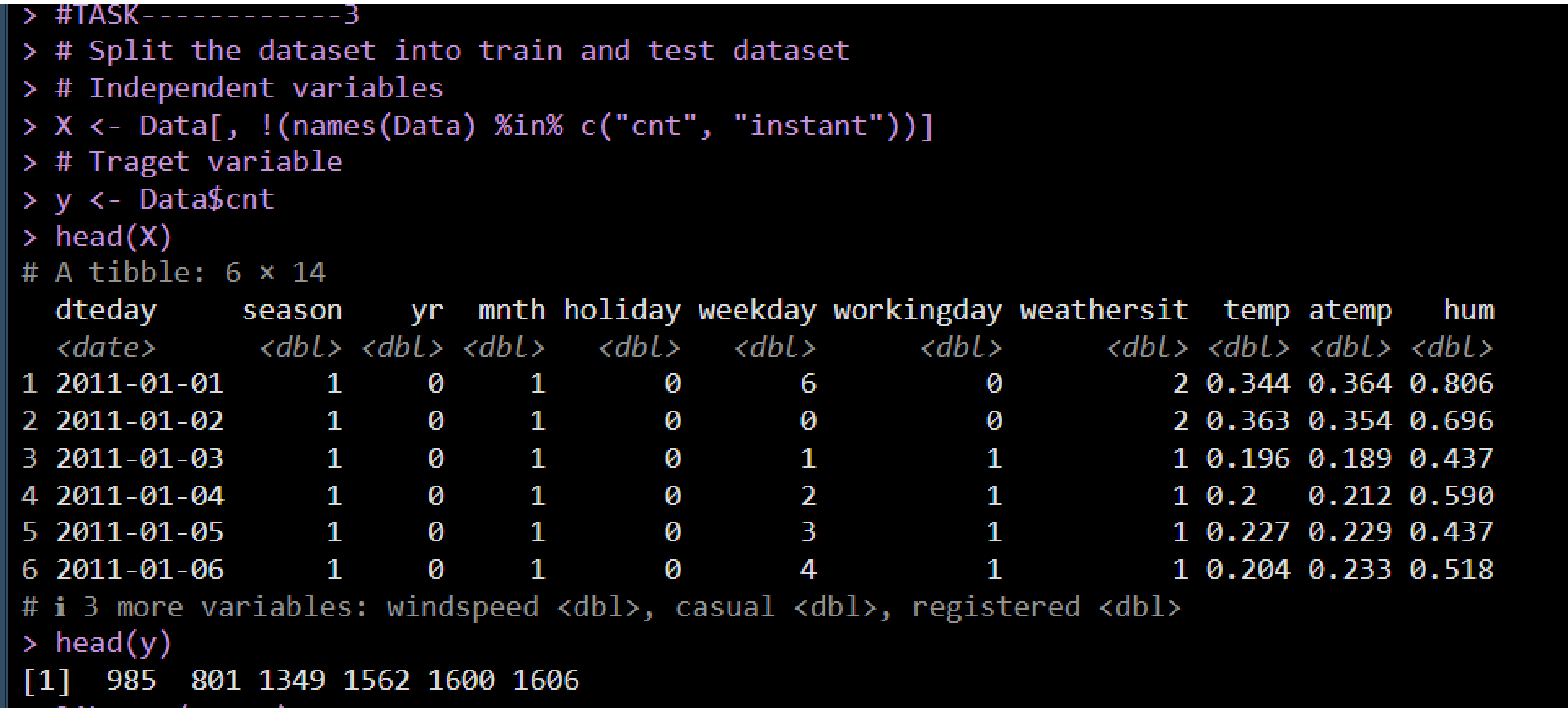
```
> Data$dteday <- as.Date(Data$dteday)
> str(Data)
tibble [731 × 16] (S3: tbl_df/tbl/data.frame)
 $ instant   : num [1:731] 1 2 3 4 5 6 7 8 9 10 ...
 $ dteday    : Date[1:731], format: "2011-01-01" "2011-01-02" ...
 $ season    : num [1:731] 1 1 1 1 1 1 1 1 1 1 ...
 $ yr        : num [1:731] 0 0 0 0 0 0 0 0 0 0 ...
 $ mnth      : num [1:731] 1 1 1 1 1 1 1 1 1 1 ...
 $ holiday   : num [1:731] 0 0 0 0 0 0 0 0 0 0 ...
 $ weekday   : num [1:731] 6 0 1 2 3 4 5 6 0 1 ...
 $ workingday: num [1:731] 0 0 1 1 1 1 1 0 0 1 ...
 $ weathersit: num [1:731] 2 2 1 1 1 1 2 2 1 1 ...
 $ temp      : num [1:731] 0.344 0.363 0.196 0.2 0.227 ...
 $ atemp     : num [1:731] 0.364 0.354 0.189 0.212 0.229 ...
 $ hum       : num [1:731] 0.806 0.696 0.437 0.59 0.437 ...
 $ windspeed : num [1:731] 0.16 0.249 0.248 0.16 0.187 ...
 $ casual    : num [1:731] 331 131 120 108 82 88 148 68 54 41 ...
 $ registered: num [1:731] 654 670 1229 1454 1518 ...
 $ cnt       : num [1:731] 985 801 1349 1562 1600 ...
> #Carry out the missing value analysis
> missing_values <- colSums(is.na(Data))
> print(missing_values)## As per the missing_values summary there is no missing values in dataset
   instant     dteday     season         yr       mnth    holiday    weekday workingday weathersit 
         0          0          0          0          0          0          0          0          0 
      temp      atemp        hum  windspeed     casual registered        cnt 
         0          0          0          0          0          0          0 
> |
```


Monthly Distribution of Bikes Rented
Total Number of Bikes Rented
3e+05
2e+05
1e+05
0e+00
0.0
2.5
5.0
7.5
10.0
12.5
Month
factor(mnth)
1
2
3
4
5
6
7
8
9
10
11
12

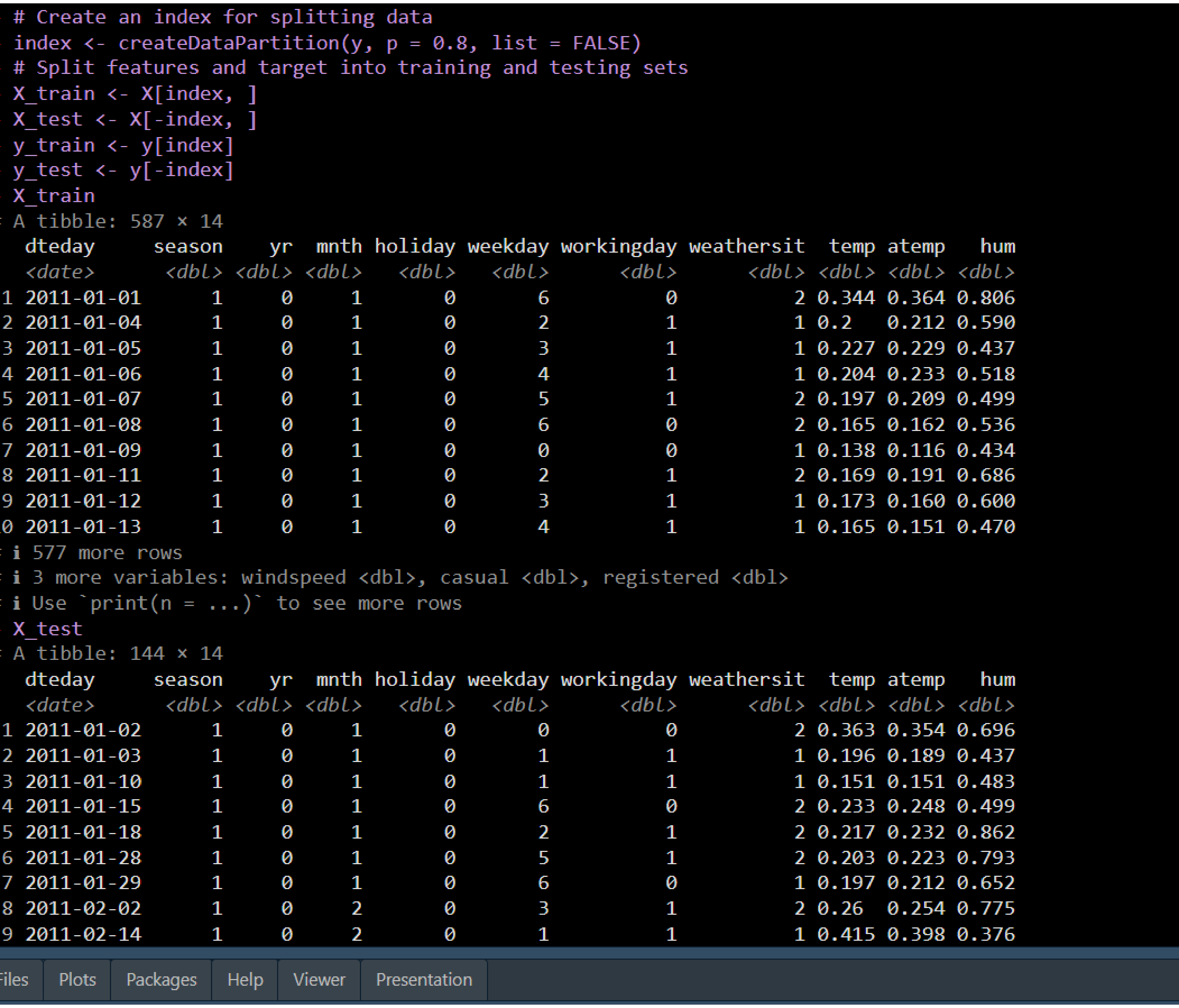


Yearly Distribution of Bikes Rented
Total Number of Bikes Rented
2000000
1500000
1000000
500000
0
-0.5
0.0
0.5
1.0
1.5
Year
factor(yr)
0
1

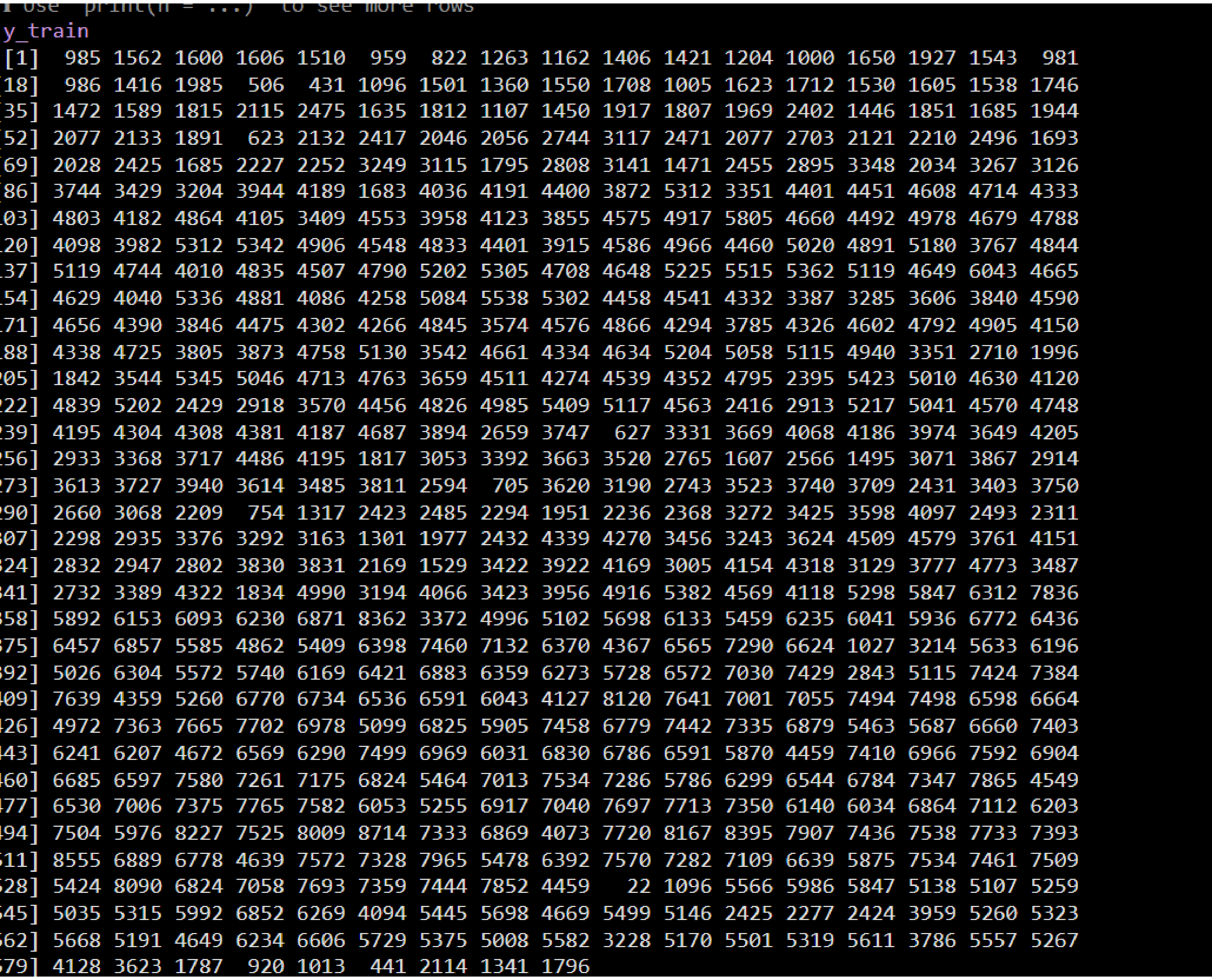

Boxplot for Outliers' Analysis
Total Number of Bikes Rented
7500
5000
2500
0
0
1
Year



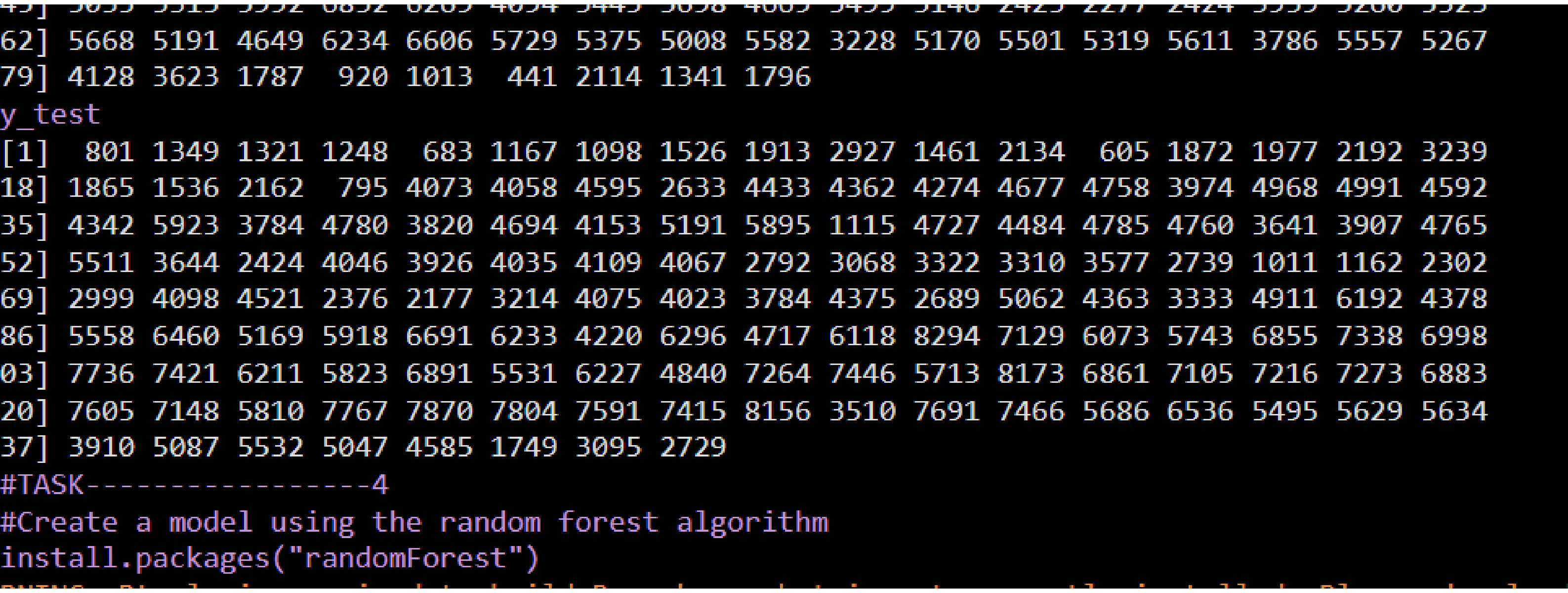
```
> #TASK------------3
> # Split the dataset into train and test dataset
> # Independent variables
> X <- Data[, !(names(Data) %in% c("cnt", "instant"))]
> # Traget variable
> y <- Data$cnt
> head(X)
# A tibble: 6 × 14
  dteday     season    yr  mnth holiday weekday workingday weathersit  temp atemp   hum
  <date>      <dbl> <dbl> <dbl>   <dbl>   <dbl>      <dbl>      <dbl> <dbl> <dbl> <dbl>
1 2011-01-01      1     0     1       0       6          0          2 0.344 0.364 0.806
2 2011-01-02      1     0     1       0       0          0          2 0.363 0.354 0.696
3 2011-01-03      1     0     1       0       1          1          1 0.196 0.189 0.437
4 2011-01-04      1     0     1       0       2          1          1 0.2   0.212 0.590
5 2011-01-05      1     0     1       0       3          1          1 0.227 0.229 0.437
6 2011-01-06      1     0     1       0       4          1          1 0.204 0.233 0.518
# ℹ 3 more variables: windspeed <dbl>, casual <dbl>, registered <dbl>
> head(y)
[1]  985  801 1349 1562 1600 1606
```



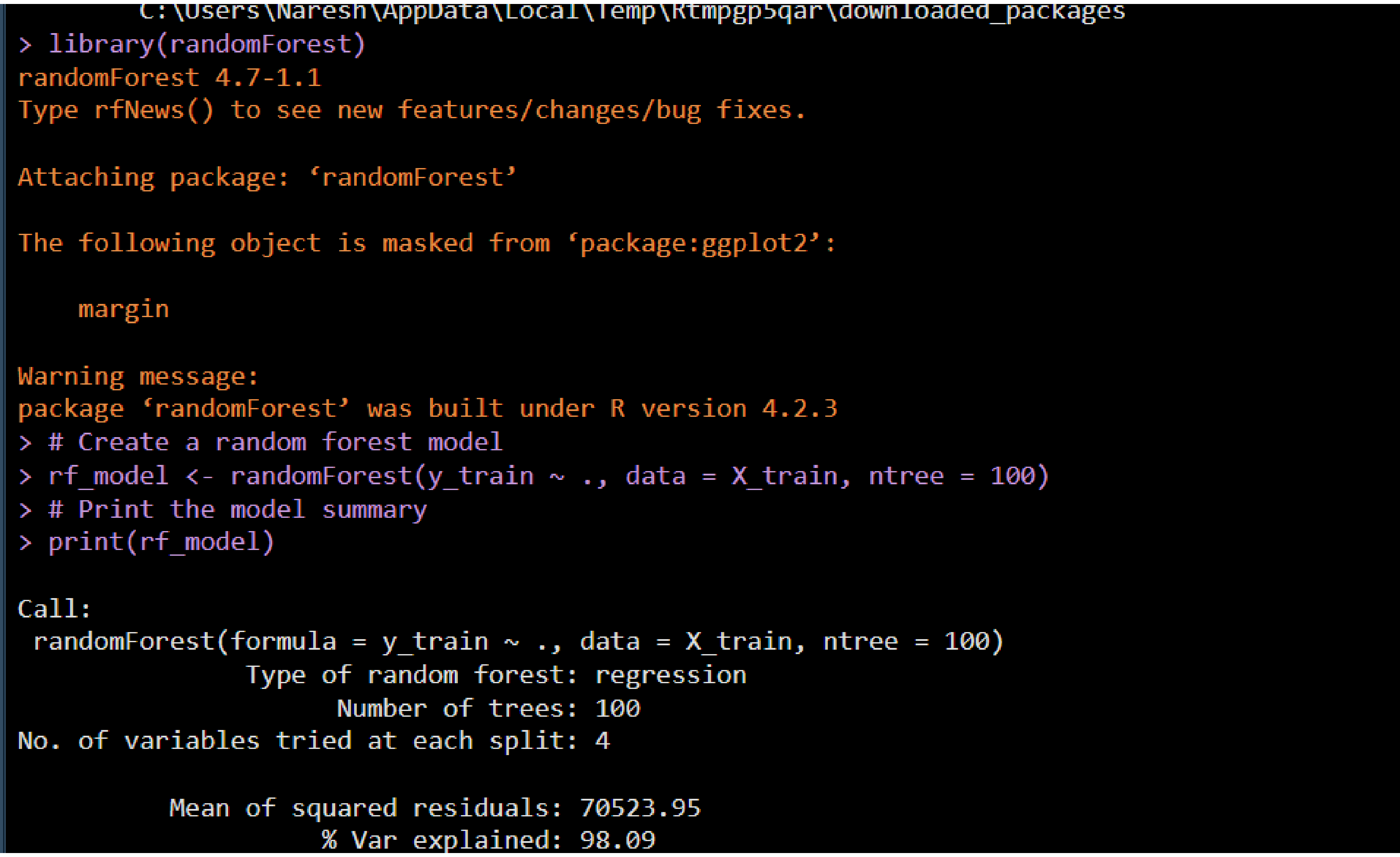
```
# Create an index for splitting data
index <- createDataPartition(y, p = 0.8, list = FALSE)
# Split features and target into training and testing sets
X_train <- X[index, ]
X_test <- X[-index, ]
y_train <- y[index]
y_test <- y[-index]
X_train
A tibble: 587 × 14
  dteday     season    yr  mnth holiday weekday workingday weathersit  temp atemp   hum
  <date>      <dbl> <dbl> <dbl>   <dbl>   <dbl>      <dbl>      <dbl> <dbl> <dbl> <dbl>
1 2011-01-01      1     0     1       0       6          0          2 0.344 0.364 0.806
2 2011-01-04      1     0     1       0       2          1          1 0.2   0.212 0.590
3 2011-01-05      1     0     1       0       3          1          1 0.227 0.229 0.437
4 2011-01-06      1     0     1       0       4          1          1 0.204 0.233 0.518
5 2011-01-07      1     0     1       0       5          1          2 0.197 0.209 0.499
6 2011-01-08      1     0     1       0       6          0          2 0.165 0.162 0.536
7 2011-01-09      1     0     1       0       0          0          1 0.138 0.116 0.434
8 2011-01-11      1     0     1       0       2          1          2 0.169 0.191 0.686
9 2011-01-12      1     0     1       0       3          1          1 0.173 0.160 0.600
0 2011-01-13      1     0     1       0       4          1          1 0.165 0.151 0.470
ℹ 577 more rows
ℹ 3 more variables: windspeed <dbl>, casual <dbl>, registered <dbl>
ℹ Use `print(n = ...)` to see more rows
X_test
A tibble: 144 × 14
  dteday     season    yr  mnth holiday weekday workingday weathersit  temp atemp   hum
  <date>      <dbl> <dbl> <dbl>   <dbl>   <dbl>      <dbl>      <dbl> <dbl> <dbl> <dbl>
1 2011-01-02      1     0     1       0       0          0          2 0.363 0.354 0.696
2 2011-01-03      1     0     1       0       1          1          1 0.196 0.189 0.437
3 2011-01-10      1     0     1       0       1          1          1 0.151 0.151 0.483
4 2011-01-15      1     0     1       0       6          0          2 0.233 0.248 0.499
5 2011-01-18      1     0     1       0       2          1          2 0.217 0.232 0.862
6 2011-01-28      1     0     1       0       5          1          2 0.203 0.223 0.793
7 2011-01-29      1     0     1       0       6          0          1 0.197 0.212 0.652
8 2011-02-02      1     0     2       0       3          1          2 0.26  0.254 0.775
9 2011-02-14      1     0     2       0       1          1          1 0.415 0.398 0.376
```



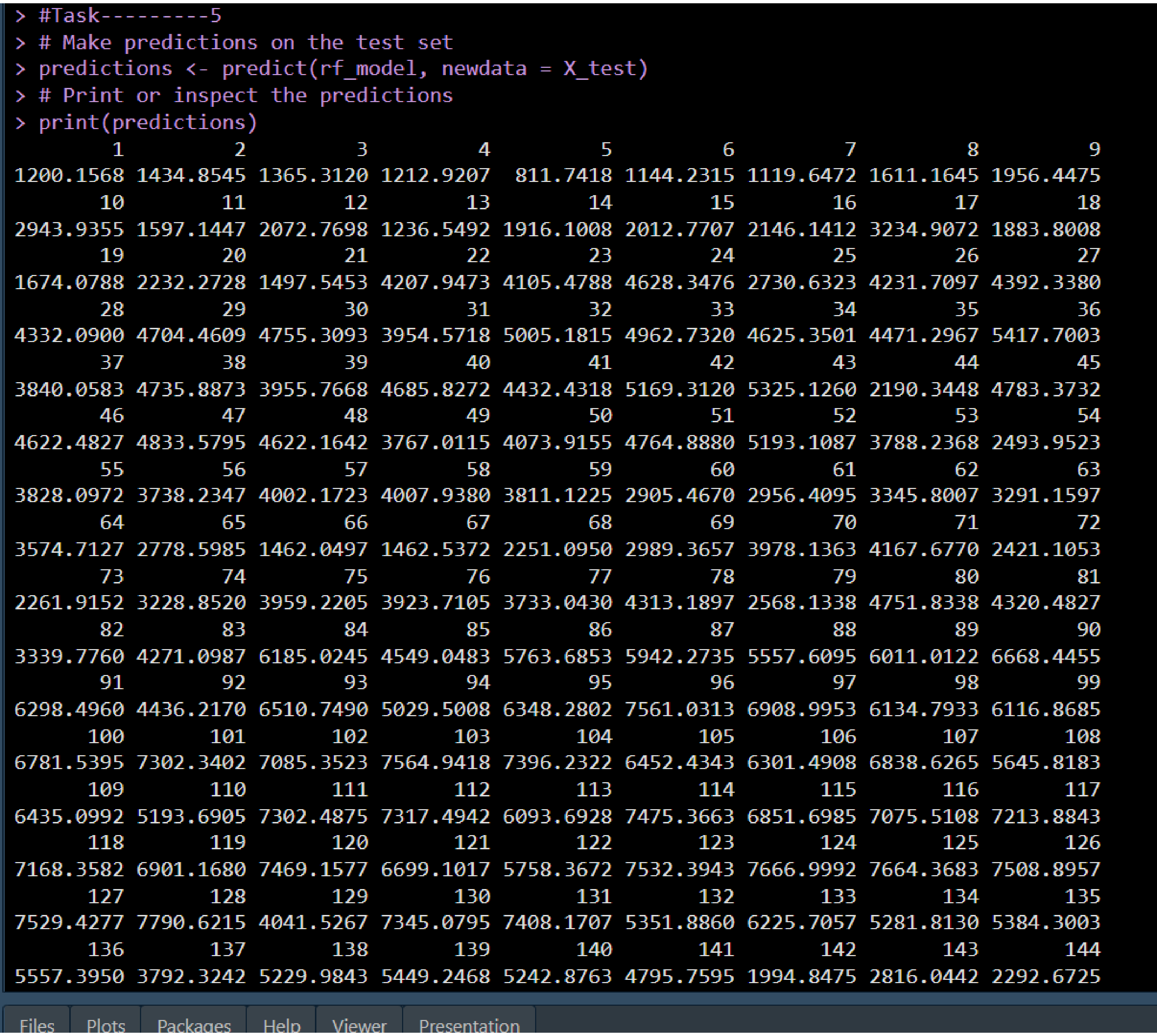
```
ℹ Use `print(n = ...)` to see more rows
y_train
 [1]  985 1562 1600 1606 1510  959  822 1263 1162 1406 1421 1204 1000 1650 1927 1543  981
[18]  986 1416 1985  506  431 1096 1501 1360 1550 1708 1005 1623 1712 1530 1605 1538 1746
[35] 1472 1589 1815 2115 2475 1635 1812 1107 1450 1917 1807 1969 2402 1446 1851 1685 1944
[52] 2077 2133 1891  623 2132 2417 2046 2056 2744 3117 2471 2077 2703 2121 2210 2496 1693
[69] 2028 2425 1685 2227 2252 3249 3115 1795 2808 3141 1471 2455 2895 3348 2034 3267 3126
[86] 3744 3429 3204 3944 4189 1683 4036 4191 4400 3872 5312 3351 4401 4451 4608 4714 4333
03] 4803 4182 4864 4105 3409 4553 3958 4123 3855 4575 4917 5805 4660 4492 4978 4679 4788
20] 4098 3982 5312 5342 4906 4548 4833 4401 3915 4586 4966 4460 5020 4891 5180 3767 4844
37] 5119 4744 4010 4835 4507 4790 5202 5305 4708 4648 5225 5515 5362 5119 4649 6043 4665
54] 4629 4040 5336 4881 4086 4258 5084 5538 5302 4458 4541 4332 3387 3285 3606 3840 4590
71] 4656 4390 3846 4475 4302 4266 4845 3574 4576 4866 4294 3785 4326 4602 4792 4905 4150
88] 4338 4725 3805 3873 4758 5130 3542 4661 4334 4634 5204 5058 5115 4940 3351 2710 1996
05] 1842 3544 5345 5046 4713 4763 3659 4511 4274 4539 4352 4795 2395 5423 5010 4630 4120
22] 4839 5202 2429 2918 3570 4456 4826 4985 5409 5117 4563 2416 2913 5217 5041 4570 4748
39] 4195 4304 4308 4381 4187 4687 3894 2659 3747  627 3331 3669 4068 4186 3974 3649 4205
56] 2933 3368 3717 4486 4195 1817 3053 3392 3663 3520 2765 1607 2566 1495 3071 3867 2914
73] 3613 3727 3940 3614 3485 3811 2594  705 3620 3190 2743 3523 3740 3709 2431 3403 3750
90] 2660 3068 2209  754 1317 2423 2485 2294 1951 2236 2368 3272 3425 3598 4097 2493 2311
07] 2298 2935 3376 3292 3163 1301 1977 2432 4339 4270 3456 3243 3624 4509 4579 3761 4151
24] 2832 2947 2802 3830 3831 2169 1529 3422 3922 4169 3005 4154 4318 3129 3777 4773 3487
41] 2732 3389 4322 1834 4990 3194 4066 3423 3956 4916 5382 4569 4118 5298 5847 6312 7836
58] 5892 6153 6093 6230 6871 8362 3372 4996 5102 5698 6133 5459 6235 6041 5936 6772 6436
75] 6457 6857 5585 4862 5409 6398 7460 7132 6370 4367 6565 7290 6624 1027 3214 5633 6196
92] 5026 6304 5572 5740 6169 6421 6883 6359 6273 5728 6572 7030 7429 2843 5115 7424 7384
09] 7639 4359 5260 6770 6734 6536 6591 6043 4127 8120 7641 7001 7055 7494 7498 6598 6664
26] 4972 7363 7665 7702 6978 5099 6825 5905 7458 6779 7442 7335 6879 5463 5687 6660 7403
43] 6241 6207 4672 6569 6290 7499 6969 6031 6830 6786 6591 5870 4459 7410 6966 7592 6904
60] 6685 6597 7580 7261 7175 6824 5464 7013 7534 7286 5786 6299 6544 6784 7347 7865 4549
77] 6530 7006 7375 7765 7582 6053 5255 6917 7040 7697 7713 7350 6140 6034 6864 7112 6203
94] 7504 5976 8227 7525 8009 8714 7333 6869 4073 7720 8167 8395 7907 7436 7538 7733 7393
11] 8555 6889 6778 4639 7572 7328 7965 5478 6392 7570 7282 7109 6639 5875 7534 7461 7509
28] 5424 8090 6824 7058 7693 7359 7444 7852 4459   22 1096 5566 5986 5847 5138 5107 5259
45] 5035 5315 5992 6852 6269 4094 5445 5698 4669 5499 5146 2425 2277 2424 3959 5260 5323
62] 5668 5191 4649 6234 6606 5729 5375 5008 5582 3228 5170 5501 5319 5611 3786 5557 5267
79] 4128 3623 1787  920 1013  441 2114 1341 1796
```



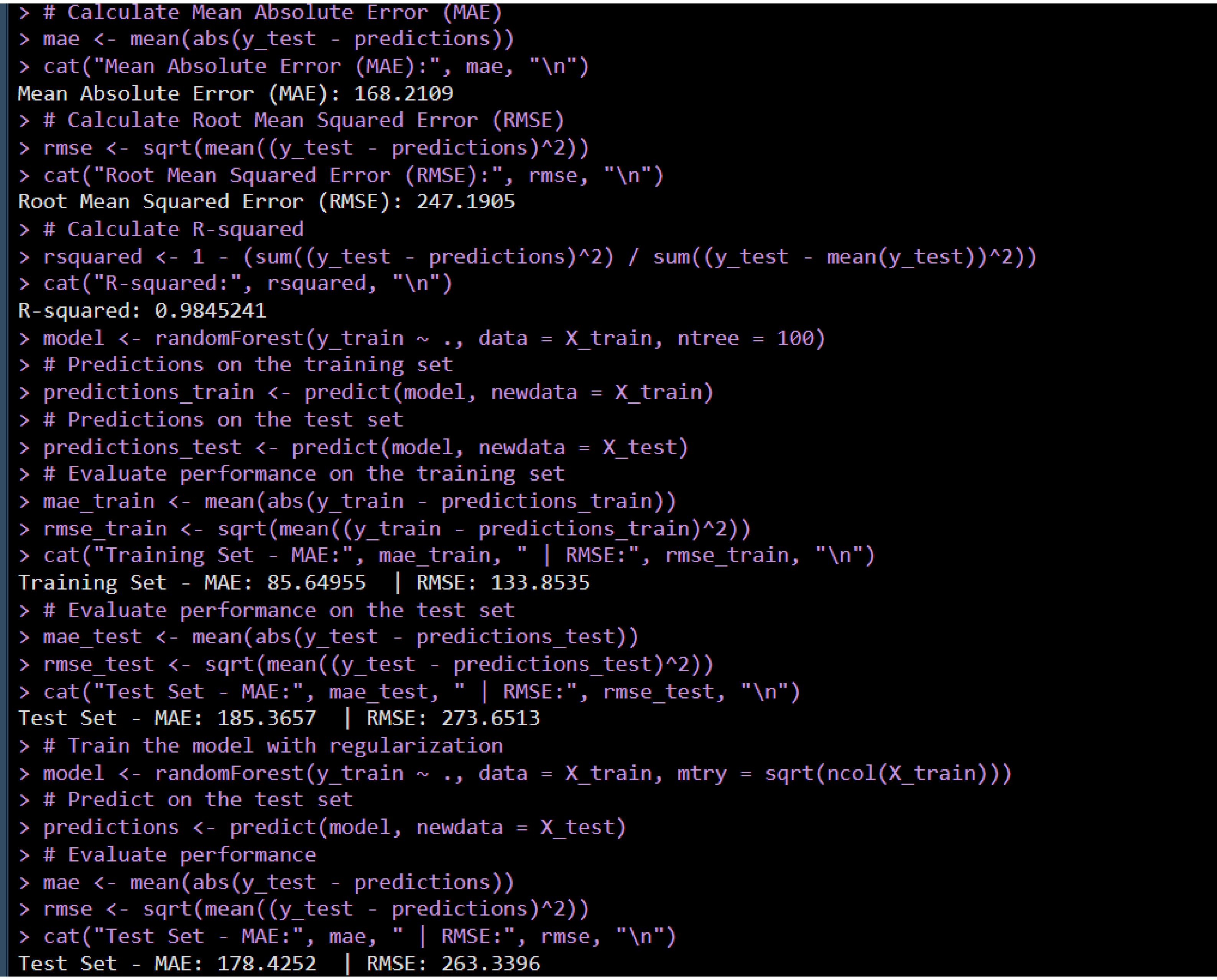
```
62] 5668 5191 4649 6234 6606 5729 5375 5008 5582 3228 5170 5501 5319 5611 3786 5557 5267
79] 4128 3623 1787  920 1013  441 2114 1341 1796
y_test
[1]  801 1349 1321 1248  683 1167 1098 1526 1913 2927 1461 2134  605 1872 1977 2192 3239
18] 1865 1536 2162  795 4073 4058 4595 2633 4433 4362 4274 4677 4758 3974 4968 4991 4592
35] 4342 5923 3784 4780 3820 4694 4153 5191 5895 1115 4727 4484 4785 4760 3641 3907 4765
52] 5511 3644 2424 4046 3926 4035 4109 4067 2792 3068 3322 3310 3577 2739 1011 1162 2302
69] 2999 4098 4521 2376 2177 3214 4075 4023 3784 4375 2689 5062 4363 3333 4911 6192 4378
86] 5558 6460 5169 5918 6691 6233 4220 6296 4717 6118 8294 7129 6073 5743 6855 7338 6998
03] 7736 7421 6211 5823 6891 5531 6227 4840 7264 7446 5713 8173 6861 7105 7216 7273 6883
20] 7605 7148 5810 7767 7870 7804 7591 7415 8156 3510 7691 7466 5686 6536 5495 5629 5634
37] 3910 5087 5532 5047 4585 1749 3095 2729
#TASK-----------------4
#Create a model using the random forest algorithm
install.packages("randomForest")
```



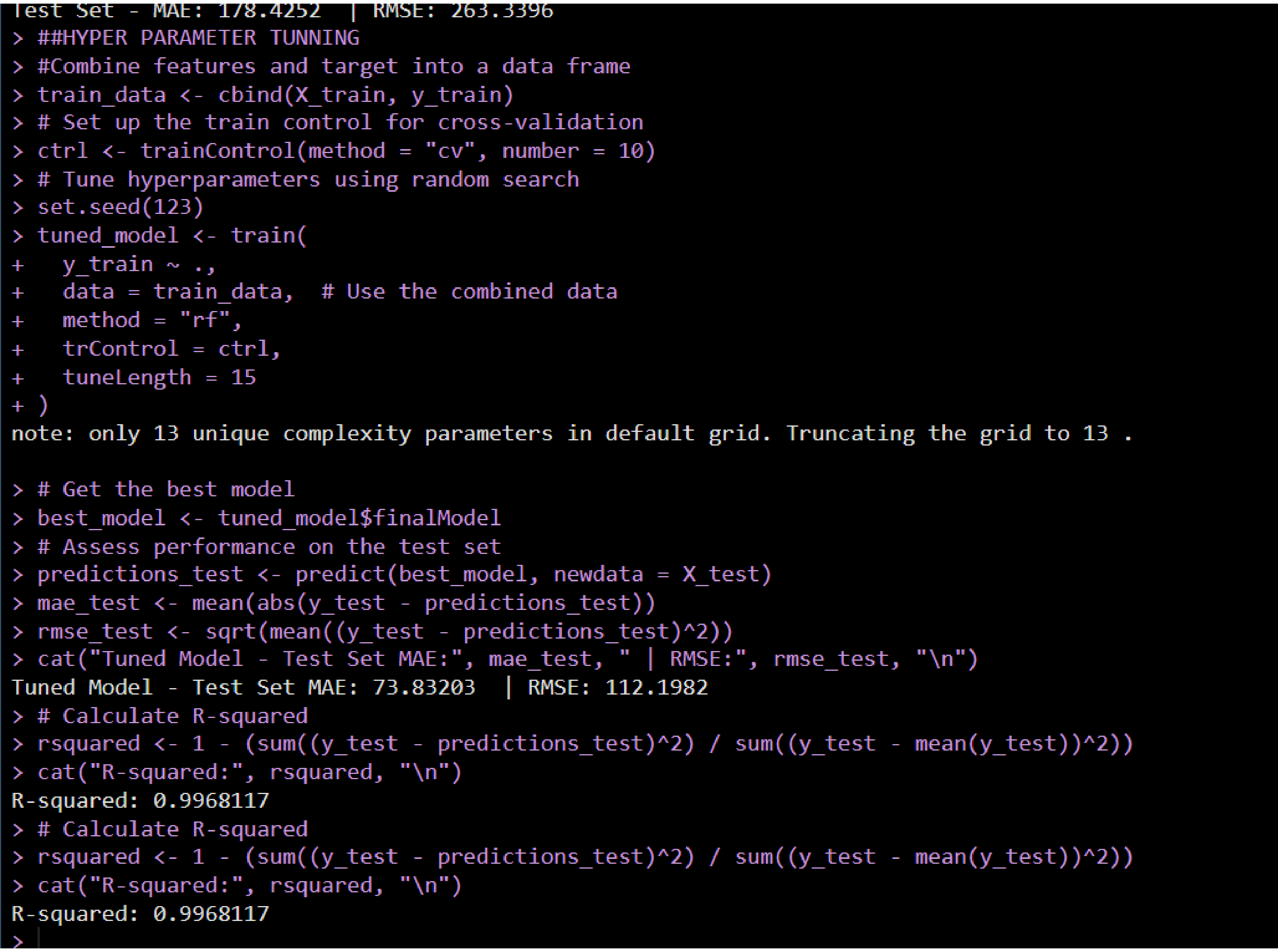
```
        C:\Users\Naresh\AppData\Local\Temp\Rtmpgp5qar\downloaded_packages
> library(randomForest)
randomForest 4.7-1.1
Type rfNews() to see new features/changes/bug fixes.

Attaching package: ‘randomForest’

The following object is masked from ‘package:ggplot2’:

    margin

Warning message:
package ‘randomForest’ was built under R version 4.2.3
> # Create a random forest model
> rf_model <- randomForest(y_train ~ ., data = X_train, ntree = 100)
> # Print the model summary
> print(rf_model)

Call:
 randomForest(formula = y_train ~ ., data = X_train, ntree = 100)
               Type of random forest: regression
                     Number of trees: 100
No. of variables tried at each split: 4

          Mean of squared residuals: 70523.95
                    % Var explained: 98.09
```



```
> #Task---------5
> # Make predictions on the test set
> predictions <- predict(rf_model, newdata = X_test)
> # Print or inspect the predictions
> print(predictions)
        1         2         3         4         5         6         7         8         9 
1200.1568 1434.8545 1365.3120 1212.9207  811.7418 1144.2315 1119.6472 1611.1645 1956.4475 
       10        11        12        13        14        15        16        17        18 
2943.9355 1597.1447 2072.7698 1236.5492 1916.1008 2012.7707 2146.1412 3234.9072 1883.8008 
       19        20        21        22        23        24        25        26        27 
1674.0788 2232.2728 1497.5453 4207.9473 4105.4788 4628.3476 2730.6323 4231.7097 4392.3380 
       28        29        30        31        32        33        34        35        36 
4332.0900 4704.4609 4755.3093 3954.5718 5005.1815 4962.7320 4625.3501 4471.2967 5417.7003 
       37        38        39        40        41        42        43        44        45 
3840.0583 4735.8873 3955.7668 4685.8272 4432.4318 5169.3120 5325.1260 2190.3448 4783.3732 
       46        47        48        49        50        51        52        53        54 
4622.4827 4833.5795 4622.1642 3767.0115 4073.9155 4764.8880 5193.1087 3788.2368 2493.9523 
       55        56        57        58        59        60        61        62        63 
3828.0972 3738.2347 4002.1723 4007.9380 3811.1225 2905.4670 2956.4095 3345.8007 3291.1597 
       64        65        66        67        68        69        70        71        72 
3574.7127 2778.5985 1462.0497 1462.5372 2251.0950 2989.3657 3978.1363 4167.6770 2421.1053 
       73        74        75        76        77        78        79        80        81 
2261.9152 3228.8520 3959.2205 3923.7105 3733.0430 4313.1897 2568.1338 4751.8338 4320.4827 
       82        83        84        85        86        87        88        89        90 
3339.7760 4271.0987 6185.0245 4549.0483 5763.6853 5942.2735 5557.6095 6011.0122 6668.4455 
       91        92        93        94        95        96        97        98        99 
6298.4960 4436.2170 6510.7490 5029.5008 6348.2802 7561.0313 6908.9953 6134.7933 6116.8685 
      100       101       102       103       104       105       106       107       108 
6781.5395 7302.3402 7085.3523 7564.9418 7396.2322 6452.4343 6301.4908 6838.6265 5645.8183 
      109       110       111       112       113       114       115       116       117 
6435.0992 5193.6905 7302.4875 7317.4942 6093.6928 7475.3663 6851.6985 7075.5108 7213.8843 
      118       119       120       121       122       123       124       125       126 
7168.3582 6901.1680 7469.1577 6699.1017 5758.3672 7532.3943 7666.9992 7664.3683 7508.8957 
      127       128       129       130       131       132       133       134       135 
7529.4277 7790.6215 4041.5267 7345.0795 7408.1707 5351.8860 6225.7057 5281.8130 5384.3003 
      136       137       138       139       140       141       142       143       144 
5557.3950 3792.3242 5229.9843 5449.2468 5242.8763 4795.7595 1994.8475 2816.0442 2292.6725 
```



```
> # Calculate Mean Absolute Error (MAE)
> mae <- mean(abs(y_test - predictions))
> cat("Mean Absolute Error (MAE):", mae, "\n")
Mean Absolute Error (MAE): 168.2109
> # Calculate Root Mean Squared Error (RMSE)
> rmse <- sqrt(mean((y_test - predictions)^2))
> cat("Root Mean Squared Error (RMSE):", rmse, "\n")
Root Mean Squared Error (RMSE): 247.1905
> # Calculate R-squared
> rsquared <- 1 - (sum((y_test - predictions)^2) / sum((y_test - mean(y_test))^2))
> cat("R-squared:", rsquared, "\n")
R-squared: 0.9845241
> model <- randomForest(y_train ~ ., data = X_train, ntree = 100)
> # Predictions on the training set
> predictions_train <- predict(model, newdata = X_train)
> # Predictions on the test set
> predictions_test <- predict(model, newdata = X_test)
> # Evaluate performance on the training set
> mae_train <- mean(abs(y_train - predictions_train))
> rmse_train <- sqrt(mean((y_train - predictions_train)^2))
> cat("Training Set - MAE:", mae_train, " | RMSE:", rmse_train, "\n")
Training Set - MAE: 85.64955  | RMSE: 133.8535
> # Evaluate performance on the test set
> mae_test <- mean(abs(y_test - predictions_test))
> rmse_test <- sqrt(mean((y_test - predictions_test)^2))
> cat("Test Set - MAE:", mae_test, " | RMSE:", rmse_test, "\n")
Test Set - MAE: 185.3657  | RMSE: 273.6513
> # Train the model with regularization
> model <- randomForest(y_train ~ ., data = X_train, mtry = sqrt(ncol(X_train)))
> # Predict on the test set
> predictions <- predict(model, newdata = X_test)
> # Evaluate performance
> mae <- mean(abs(y_test - predictions))
> rmse <- sqrt(mean((y_test - predictions)^2))
> cat("Test Set - MAE:", mae, " | RMSE:", rmse, "\n")
Test Set - MAE: 178.4252  | RMSE: 263.3396
```



```
Test Set - MAE: 178.4252  | RMSE: 263.3396
> ##HYPER PARAMETER TUNNING
> #Combine features and target into a data frame
> train_data <- cbind(X_train, y_train)
> # Set up the train control for cross-validation
> ctrl <- trainControl(method = "cv", number = 10)
> # Tune hyperparameters using random search
> set.seed(123)
> tuned_model <- train(
+   y_train ~ .,
+   data = train_data,  # Use the combined data
+   method = "rf",
+   trControl = ctrl,
+   tuneLength = 15
+ )
note: only 13 unique complexity parameters in default grid. Truncating the grid to 13 .

> # Get the best model
> best_model <- tuned_model$finalModel
> # Assess performance on the test set
> predictions_test <- predict(best_model, newdata = X_test)
> mae_test <- mean(abs(y_test - predictions_test))
> rmse_test <- sqrt(mean((y_test - predictions_test)^2))
> cat("Tuned Model - Test Set MAE:", mae_test, " | RMSE:", rmse_test, "\n")
Tuned Model - Test Set MAE: 73.83203  | RMSE: 112.1982
> # Calculate R-squared
> rsquared <- 1 - (sum((y_test - predictions_test)^2) / sum((y_test - mean(y_test))^2))
> cat("R-squared:", rsquared, "\n")
R-squared: 0.9968117
> # Calculate R-squared
> rsquared <- 1 - (sum((y_test - predictions_test)^2) / sum((y_test - mean(y_test))^2))
> cat("R-squared:", rsquared, "\n")
R-squared: 0.9968117
> 
```