Prediction of Online Sales using Xgboost

Input data

```
data=read.csv('F:\\Datasets\\Online Sales Prediction\\train.csv',header=T)
```

Characteristics of the data

```
dim(data)
```

```
## [1] 1017209       9
```

```
head(data)
```

```
##   Store DayOfWeek       Date Sales Customers Open Promo StateHoliday
## 1     1         5 2015-07-31  5263       555    1     1            0
## 2     2         5 2015-07-31  6064       625    1     1            0
## 3     3         5 2015-07-31  8314       821    1     1            0
## 4     4         5 2015-07-31 13995      1498    1     1            0
## 5     5         5 2015-07-31  4822       559    1     1            0
## 6     6         5 2015-07-31  5651       589    1     1            0
##   SchoolHoliday
## 1             1
## 2             1
## 3             1
## 4             1
## 5             1
## 6             1
```

```
str(data)
```

```
## 'data.frame':    1017209 obs. of  9 variables:
##  $ Store        : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ DayOfWeek    : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ Date         : Factor w/ 942 levels "2013-01-01","2013-01-02",..: 942 942 942 9
##  $ Sales        : int  5263 6064 8314 13995 4822 5651 15344 8492 8565 7185 ...
##  $ Customers    : int  555 625 821 1498 559 589 1414 833 687 681 ...
##  $ Open         : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Promo        : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ StateHoliday : Factor w/ 4 levels "0","a","b","c": 1 1 1 1 1 1 1 1 1 1 ...
##  $ SchoolHoliday: int  1 1 1 1 1 1 1 1 1 1 ...
```

```
summary(data)
```

```
##      Store          DayOfWeek             Date             Sales       
##  Min.   :   1.0   Min.   :1.000   2013-01-02:   1115   Min.   :    0  
##  1st Qu.: 280.0   1st Qu.:2.000   2013-01-03:   1115   1st Qu.: 3727  
##  Median : 558.0   Median :4.000   2013-01-04:   1115   Median : 5744  
##  Mean   : 558.4   Mean   :3.998   2013-01-05:   1115   Mean   : 5774  
##  3rd Qu.: 838.0   3rd Qu.:6.000   2013-01-06:   1115   3rd Qu.: 7856  
##  Max.   :1115.0   Max.   :7.000   2013-01-07:   1115   Max.   :41551  
##                                   (Other)   :1010519                  
##    Customers           Open            Promo        StateHoliday
##  Min.   :   0.0   Min.   :0.0000   Min.   :0.0000   0:986159    
##  1st Qu.: 405.0   1st Qu.:1.0000   1st Qu.:0.0000   a: 20260    
##  Median : 609.0   Median :1.0000   Median :0.0000   b:  6690    
##  Mean   : 633.1   Mean   :0.8301   Mean   :0.3815   c:  4100    
##  3rd Qu.: 837.0   3rd Qu.:1.0000   3rd Qu.:1.0000               
##  Max.   :7388.0   Max.   :1.0000   Max.   :1.0000               
##                                                                 
##  SchoolHoliday  
##  Min.   :0.0000  
##  1st Qu.:0.0000  
##  Median :0.0000  
##  Mean   :0.1786  
##  3rd Qu.:0.0000  
##  Max.   :1.0000  
##
```

Deleting Store and Date Variables

```
data$Store=NULL
data$Date=NULL
```

Assigning proper data type to variables

```
data$Sales <- as.numeric(data$Sales)
data$DayOfWeek <- as.factor(data$DayOfWeek)
data$Customers <- as.numeric(data$Customers)
```

Checking for NA

```
sum(is.na(data))
```

```
## [1] 0
```

Total occurences of zero Sales (Target)

```
sum(data$Sales==0)
```

```
## [1] 172871
```

Deleting zero values

```
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
```

```
data <- data%>%filter(Sales>0)
dim(data)
```

```
## [1] 844338      7
```

Data Manipulation

```
# The table of average daily sales is
 data%>%group_by(DayOfWeek)%>%summarise(Daily_Average=mean(Sales))
```

```
## # A tibble: 7 x 2
##   DayOfWeek Daily_Average
##   <fct>             <dbl>
## 1 1                 8216.
## 2 2                 7088.
## 3 3                 6729.
## 4 4                 6768.
## 5 5                 7073.
## 6 6                 5875.
## 7 7                 8225.
```

The average sales by School Holiday is

```
# 0 - No School Holiday ; 1 - School Holiday
 data%>%group_by(SchoolHoliday)%>%summarise(avg_on_holiday=mean(Sales))
```

```
## # A tibble: 2 x 2
##   SchoolHoliday avg_on_holiday
##           <int>          <dbl>
## 1             0          6897.
## 2             1          7201.
```

The average daily sales and the median of customers served by Promo code

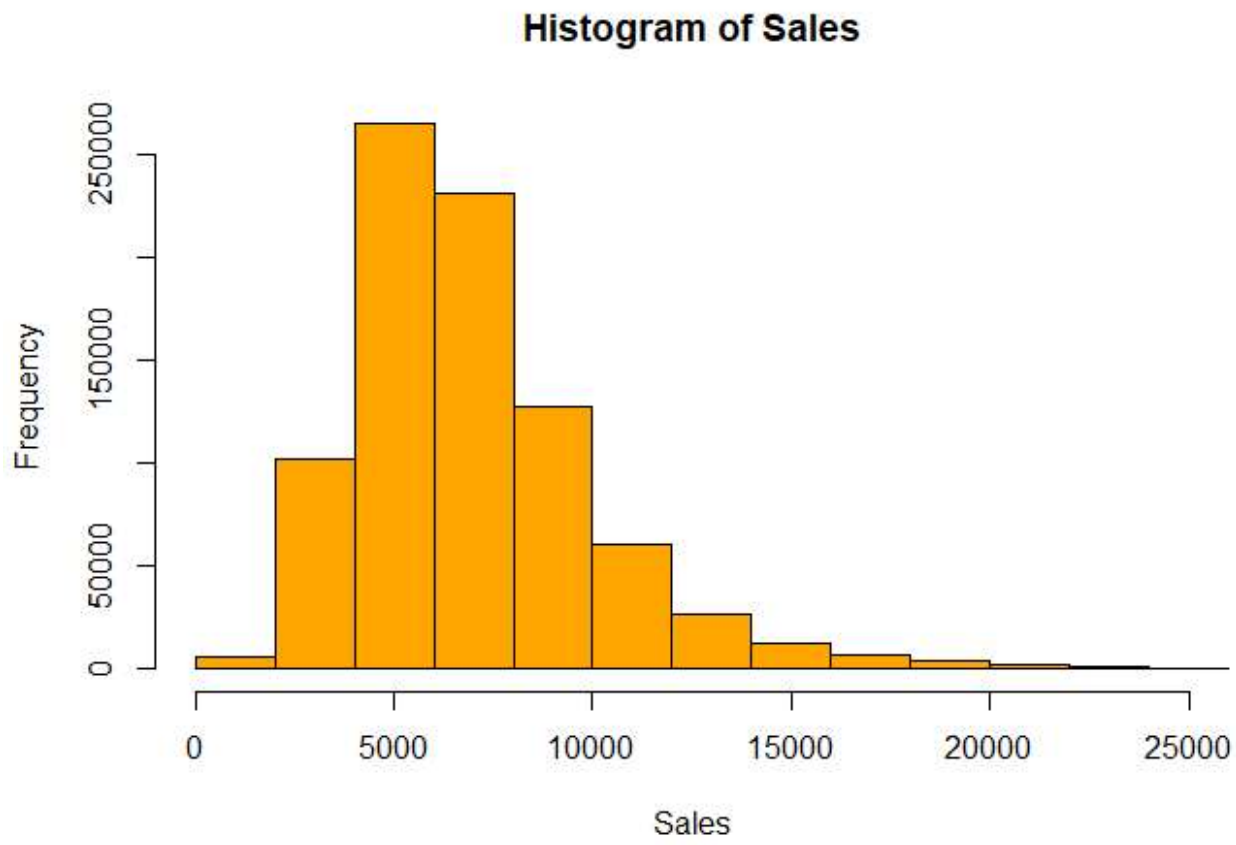
```
data%>%group_by(Promo,DayOfWeek)%>%summarise(max_sales=max(Sales),median_of_customers
```

```
## # A tibble: 12 x 4
## # Groups:   Promo [2]
##    Promo DayOfWeek max_sales median_of_customers
##    <int> <fct>         <dbl>               <dbl>
##  1     0 1             41551                 651
##  2     1 5             38722                 739
##  3     0 4             38367                 611
##  4     1 1             38037                 858
##  5     0 7             37376                1262
##  6     1 4             34814                 724
##  7     1 2             34692                 753
##  8     0 5             33934                 653
##  9     1 3             33151                 714
## 10     0 2             31930                 604
## 11     0 6             31683                 573
## 12     0 3             26818                 597
```

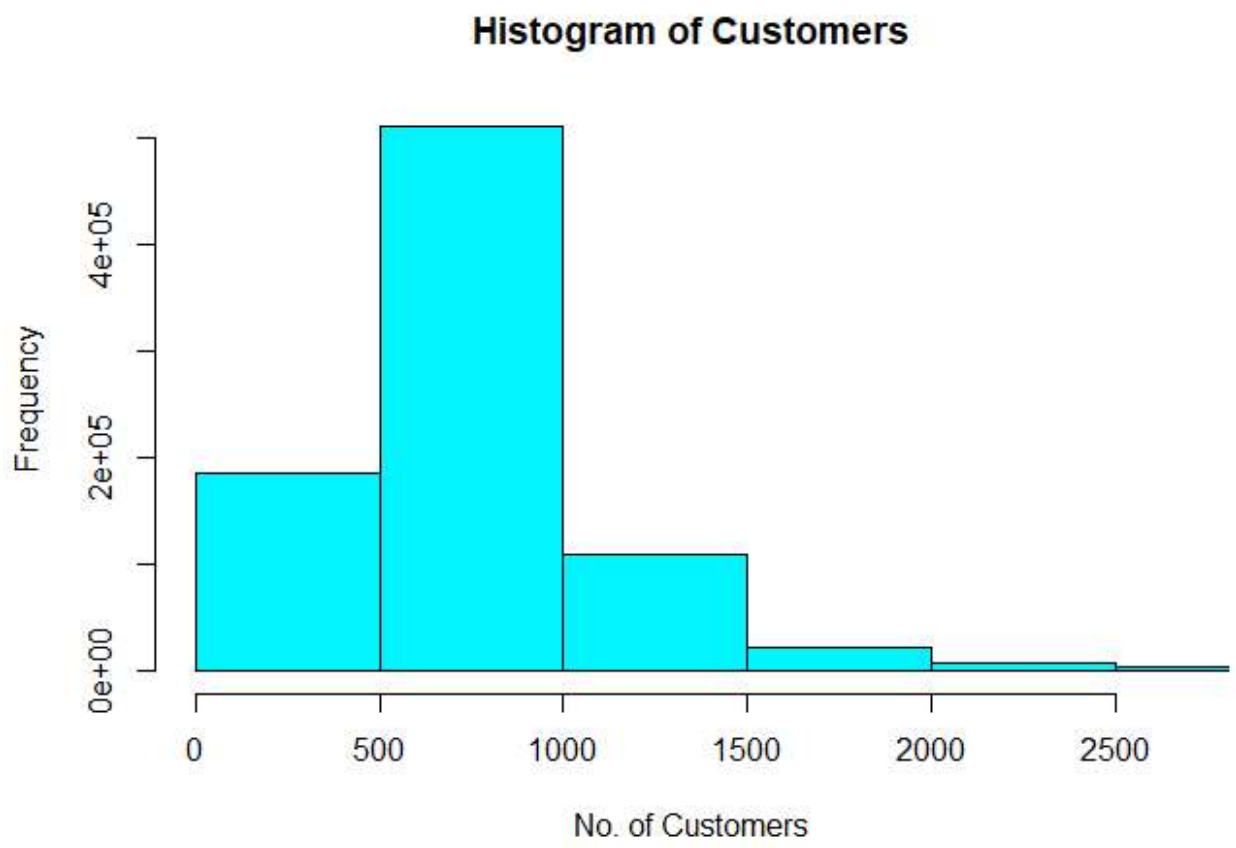
Data Visualization

Histograms for Sales and Customers

```
hist(data$Sales,xlim = c(0,25000),col='orange',main='Histogram of Sales',xlab='Sales'
     ylab='Frequency')
```




```
hist(data$Customers,xlim=c(0,2700),col='turquoise1',main='Histogram of Customers',
     xlab='No. of Customers',ylab='Frequency')
```



Scatter plot of Sales vs Customers

```
library(ggplot2)
ggplot(data)+geom_point(mapping=aes(x=Sales,y=Customers),position = 'jitter',colour='
  theme_classic()+labs(x='Sales',y='Customers',title="Scatter plot of  Customers vs S
```

Correlation

```
cor(data$Customers,data$Sales)
```

```
## [1] 0.8235517
```

Data Partition into training and test

```
set.seed(111)
idx <- sample(2,nrow(data),prob=c(0.7,0.3),replace=T)
train <- data[idx==1,]
test <- data[idx==2,]
dim(train)
```

```
## [1] 590145      7
```

```
dim(test)
```

```
## [1] 254193      7
```

One hot encoding for training set

```
data_ohe <-as.data.frame(model.matrix(~.-1,data=train))
ohe_label <- data_ohe['Sales'] # Target variable - Sales
```

One hot encoding for test set

```
test_ohe <- as.data.frame(model.matrix(~.-1,data=test))
test_label <- test_ohe['Sales']
```

Xgboost Model

```
library(xgboost)
```

```
##
## Attaching package: 'xgboost'

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

```
dtrain <- xgb.DMatrix(as.matrix(data_ohe%>%select(-Sales)),label=ohe_label)
dtest <- xgb.DMatrix(as.matrix(test_ohe%>%select(-Sales)),label=test_label)
```

Training the xgboost model

```
set.seed(500)

w <- list(train=dtrain,test= dtest)

xgb_model1 <- xgb.train(data=dtrain,booster='gbtree',nrounds=800,max_depth=6,eval_met
                        eta=0.135,watchlist=w,early_stopping_rounds = 30)
```

```
## [1]  train-rmse:6634.682129  test-rmse:6626.620117
## Multiple eval metrics are present. Will use test_rmse for early stopping.
## Will train until test_rmse hasn't improved in 30 rounds.
##
## [2]  train-rmse:5789.277344  test-rmse:5781.732910
## [3]  train-rmse:5064.992676  test-rmse:5057.577148
## [4]  train-rmse:4446.244141  test-rmse:4438.996094
## [5]  train-rmse:3919.754395  test-rmse:3912.774658
## [6]  train-rmse:3473.498047  test-rmse:3466.785645
## [7]  train-rmse:3097.755127  test-rmse:3091.212158
## [8]  train-rmse:2783.359863  test-rmse:2777.112061
## [9]  train-rmse:2522.162842  test-rmse:2516.188477
## [10] train-rmse:2307.276611  test-rmse:2301.458740
## [11] train-rmse:2132.010254  test-rmse:2126.546387
## [12] train-rmse:1990.651245  test-rmse:1985.732056
## [13] train-rmse:1877.574097  test-rmse:1873.045654
## [14] train-rmse:1788.107666  test-rmse:1784.157349
## [15] train-rmse:1718.046509  test-rmse:1714.492432
## [16] train-rmse:1663.198120  test-rmse:1660.026611
## [17] train-rmse:1621.156860  test-rmse:1618.375000
## [18] train-rmse:1588.680908  test-rmse:1586.194214
## [19] train-rmse:1563.844482  test-rmse:1561.676758
## [20] train-rmse:1544.490356  test-rmse:1542.672729
## [21] train-rmse:1529.903809  test-rmse:1528.286987
## [22] train-rmse:1518.926392  test-rmse:1517.564819
## [23] train-rmse:1510.503662  test-rmse:1509.443237
## [24] train-rmse:1504.017578  test-rmse:1503.202881
## [25] train-rmse:1499.115356  test-rmse:1498.530884
## [26] train-rmse:1495.409546  test-rmse:1495.086426
## [27] train-rmse:1492.302734  test-rmse:1492.264038
## [28] train-rmse:1490.074829  test-rmse:1490.200684
## [29] train-rmse:1488.334839  test-rmse:1488.603516
## [30] train-rmse:1487.016357  test-rmse:1487.410889
## [31] train-rmse:1486.035156  test-rmse:1486.546143
## [32] train-rmse:1485.204956  test-rmse:1485.897827
## [33] train-rmse:1484.603760  test-rmse:1485.466064
## [34] train-rmse:1483.944946  test-rmse:1484.964355
```

```
## [35] train-rmse:1483.528198  test-rmse:1484.614380
## [36] train-rmse:1483.107300  test-rmse:1484.356079
## [37] train-rmse:1482.877441  test-rmse:1484.208374
## [38] train-rmse:1482.321411  test-rmse:1483.835205
## [39] train-rmse:1481.987549  test-rmse:1483.736572
## [40] train-rmse:1481.773315  test-rmse:1483.578979
## [41] train-rmse:1481.643311  test-rmse:1483.508545
## [42] train-rmse:1481.528564  test-rmse:1483.456177
## [43] train-rmse:1481.399658  test-rmse:1483.367432
## [44] train-rmse:1481.279297  test-rmse:1483.278198
## [45] train-rmse:1480.977417  test-rmse:1483.154785
## [46] train-rmse:1480.435791  test-rmse:1482.873779
## [47] train-rmse:1480.368530  test-rmse:1482.853027
## [48] train-rmse:1480.276978  test-rmse:1482.828491
## [49] train-rmse:1480.218262  test-rmse:1482.830200
## [50] train-rmse:1480.175537  test-rmse:1482.816162
## [51] train-rmse:1479.798096  test-rmse:1482.705078
## [52] train-rmse:1479.721802  test-rmse:1482.669922
## [53] train-rmse:1479.648560  test-rmse:1482.660400
## [54] train-rmse:1479.234375  test-rmse:1482.427612
## [55] train-rmse:1479.218750  test-rmse:1482.422119
## [56] train-rmse:1479.158936  test-rmse:1482.390137
## [57] train-rmse:1479.041626  test-rmse:1482.431885
## [58] train-rmse:1478.637695  test-rmse:1482.372314
## [59] train-rmse:1478.586792  test-rmse:1482.375122
## [60] train-rmse:1478.544678  test-rmse:1482.349609
## [61] train-rmse:1478.389526  test-rmse:1482.393311
## [62] train-rmse:1478.271484  test-rmse:1482.411255
## [63] train-rmse:1478.246704  test-rmse:1482.417480
## [64] train-rmse:1477.953003  test-rmse:1482.489990
## [65] train-rmse:1477.684204  test-rmse:1482.484985
## [66] train-rmse:1477.380249  test-rmse:1482.369995
## [67] train-rmse:1477.175049  test-rmse:1482.414551
## [68] train-rmse:1477.139038  test-rmse:1482.434937
## [69] train-rmse:1476.926147  test-rmse:1482.440674
## [70] train-rmse:1476.747192  test-rmse:1482.447632
## [71] train-rmse:1476.613403  test-rmse:1482.486450
## [72] train-rmse:1476.564209  test-rmse:1482.488403
## [73] train-rmse:1476.432007  test-rmse:1482.480469
## [74] train-rmse:1476.408691  test-rmse:1482.501953

## [75] train-rmse:1476.275269  test-rmse:1482.504517
## [76] train-rmse:1476.157227  test-rmse:1482.471924
## [77] train-rmse:1475.915771  test-rmse:1482.496948
## [78] train-rmse:1475.853882  test-rmse:1482.509888
## [79] train-rmse:1475.682007  test-rmse:1482.569092
## [80] train-rmse:1475.516602  test-rmse:1482.543579
## [81] train-rmse:1475.419434  test-rmse:1482.613281
## [82] train-rmse:1475.297119  test-rmse:1482.603638
## [83] train-rmse:1475.156860  test-rmse:1482.669556
## [84] train-rmse:1475.048706  test-rmse:1482.717407
## [85] train-rmse:1474.936646  test-rmse:1482.704346
```

```
## [86] train-rmse:1474.883301  test-rmse:1482.735718
## [87] train-rmse:1474.783203  test-rmse:1482.764771
## [88] train-rmse:1474.639282  test-rmse:1482.803467
## [89] train-rmse:1474.422485  test-rmse:1482.839478
## [90] train-rmse:1474.345337  test-rmse:1482.854126
## Stopping. Best iteration:
## [60] train-rmse:1478.544678  test-rmse:1482.349609
```

Model 2

```
xgb_model2 <- xgb.train(data=dtrain,booster='gbtree',nrounds=800,max_depth=6,eval_met
                 eta=0.1,watchlist=w,early_stopping_rounds = 30)
```

```
## [1]  train-rmse:6889.255859  test-rmse:6881.096191
## Multiple eval metrics are present. Will use test_rmse for early stopping.
## Will train until test_rmse hasn't improved in 30 rounds.
##
## [2]  train-rmse:6235.568359  test-rmse:6227.811035
## [3]  train-rmse:5650.816406  test-rmse:5643.126465
## [4]  train-rmse:5128.273438  test-rmse:5120.769043
## [5]  train-rmse:4662.144531  test-rmse:4654.789062
## [6]  train-rmse:4247.190918  test-rmse:4240.054688
## [7]  train-rmse:3878.384277  test-rmse:3871.440918
## [8]  train-rmse:3551.668457  test-rmse:3544.884521
## [9]  train-rmse:3263.005371  test-rmse:3256.394775
## [10] train-rmse:3008.710205  test-rmse:3002.257568
## [11] train-rmse:2785.531982  test-rmse:2779.270996
## [12] train-rmse:2590.678955  test-rmse:2584.555176
## [13] train-rmse:2420.921631  test-rmse:2415.067139
## [14] train-rmse:2274.069580  test-rmse:2268.491211
## [15] train-rmse:2147.823242  test-rmse:2142.476074
## [16] train-rmse:2039.728638  test-rmse:2034.714600
## [17] train-rmse:1947.488525  test-rmse:1942.823608
## [18] train-rmse:1869.466187  test-rmse:1865.267334
## [19] train-rmse:1803.599365  test-rmse:1799.663330
## [20] train-rmse:1748.241577  test-rmse:1744.615234
## [21] train-rmse:1702.010376  test-rmse:1698.643921
## [22] train-rmse:1663.474854  test-rmse:1660.449219
## [23] train-rmse:1631.571655  test-rmse:1628.812622
## [24] train-rmse:1605.214355  test-rmse:1602.724854
## [25] train-rmse:1583.494019  test-rmse:1581.223389
## [26] train-rmse:1565.523071  test-rmse:1563.539062
## [27] train-rmse:1550.761719  test-rmse:1548.969360
## [28] train-rmse:1538.708740  test-rmse:1537.119141
## [29] train-rmse:1528.764893  test-rmse:1527.373901
```

```
## [30] train-rmse:1520.502686  test-rmse:1519.357300
## [31] train-rmse:1513.879272  test-rmse:1512.943726
## [32] train-rmse:1508.387695  test-rmse:1507.603394
## [33] train-rmse:1503.703003  test-rmse:1503.106323
## [34] train-rmse:1500.057251  test-rmse:1499.618652
## [35] train-rmse:1497.002808  test-rmse:1496.733643
## [36] train-rmse:1494.351807  test-rmse:1494.251587
## [37] train-rmse:1492.260132  test-rmse:1492.264771
## [38] train-rmse:1490.525513  test-rmse:1490.682617
## [39] train-rmse:1489.122437  test-rmse:1489.425659
## [40] train-rmse:1487.895752  test-rmse:1488.400024
## [41] train-rmse:1486.893555  test-rmse:1487.472656
## [42] train-rmse:1486.109375  test-rmse:1486.753296
## [43] train-rmse:1485.450806  test-rmse:1486.186035
## [44] train-rmse:1484.875122  test-rmse:1485.699219
## [45] train-rmse:1484.413574  test-rmse:1485.268677
## [46] train-rmse:1484.012939  test-rmse:1484.894287
## [47] train-rmse:1483.663940  test-rmse:1484.628906
## [48] train-rmse:1483.265015  test-rmse:1484.412964
## [49] train-rmse:1483.013428  test-rmse:1484.256958
## [50] train-rmse:1482.714966  test-rmse:1483.999268
## [51] train-rmse:1482.488037  test-rmse:1483.877930
## [52] train-rmse:1482.296143  test-rmse:1483.766113
## [53] train-rmse:1482.082642  test-rmse:1483.629028
## [54] train-rmse:1481.976196  test-rmse:1483.578979
## [55] train-rmse:1481.611206  test-rmse:1483.296143
## [56] train-rmse:1481.505127  test-rmse:1483.250854
## [57] train-rmse:1481.403931  test-rmse:1483.222900
## [58] train-rmse:1481.283936  test-rmse:1483.166870
## [59] train-rmse:1481.212891  test-rmse:1483.149414
## [60] train-rmse:1481.070190  test-rmse:1483.070923
## [61] train-rmse:1480.937256  test-rmse:1483.070923
## [62] train-rmse:1480.651245  test-rmse:1483.030029
## [63] train-rmse:1480.608521  test-rmse:1483.012451

## [64] train-rmse:1480.321045  test-rmse:1482.943481
## [65] train-rmse:1480.252441  test-rmse:1482.939697
## [66] train-rmse:1480.141235  test-rmse:1482.911621
## [67] train-rmse:1480.067017  test-rmse:1482.885620
## [68] train-rmse:1479.807373  test-rmse:1482.826904
## [69] train-rmse:1479.748169  test-rmse:1482.829346
## [70] train-rmse:1479.675781  test-rmse:1482.828857
## [71] train-rmse:1479.421509  test-rmse:1482.688354
## [72] train-rmse:1479.145752  test-rmse:1482.586304
## [73] train-rmse:1479.020264  test-rmse:1482.552002
## [74] train-rmse:1478.737427  test-rmse:1482.543823
## [75] train-rmse:1478.701538  test-rmse:1482.539429
## [76] train-rmse:1478.588379  test-rmse:1482.532837
## [77] train-rmse:1478.470337  test-rmse:1482.496582
## [78] train-rmse:1478.454834  test-rmse:1482.502686
## [79] train-rmse:1478.345825  test-rmse:1482.500732
## [80] train-rmse:1478.194214  test-rmse:1482.523315
```

```
## [81] train-rmse:1478.089966  test-rmse:1482.512207
## [82] train-rmse:1477.859131  test-rmse:1482.498779
## [83] train-rmse:1477.651001  test-rmse:1482.531250
## [84] train-rmse:1477.638184  test-rmse:1482.538330
## [85] train-rmse:1477.595947  test-rmse:1482.551392
## [86] train-rmse:1477.534302  test-rmse:1482.561890
## [87] train-rmse:1477.422852  test-rmse:1482.570679
## [88] train-rmse:1477.266602  test-rmse:1482.534546
## [89] train-rmse:1477.175903  test-rmse:1482.536255
## [90] train-rmse:1477.134644  test-rmse:1482.532104
## [91] train-rmse:1477.068115  test-rmse:1482.510986
## [92] train-rmse:1477.004028  test-rmse:1482.513916
## [93] train-rmse:1476.895386  test-rmse:1482.497681
## [94] train-rmse:1476.753174  test-rmse:1482.496094
## [95] train-rmse:1476.702759  test-rmse:1482.494141
## [96] train-rmse:1476.503540  test-rmse:1482.519775
## [97] train-rmse:1476.377930  test-rmse:1482.561890
## [98] train-rmse:1476.269165  test-rmse:1482.573120
## [99] train-rmse:1476.051514  test-rmse:1482.557129
## [100]    train-rmse:1475.920410  test-rmse:1482.590210
## [101]    train-rmse:1475.915039  test-rmse:1482.597168
## [102]    train-rmse:1475.840820  test-rmse:1482.544922
## [103]    train-rmse:1475.752563  test-rmse:1482.536743
## [104]    train-rmse:1475.661499  test-rmse:1482.575684
## [105]    train-rmse:1475.550537  test-rmse:1482.611938
## [106]    train-rmse:1475.485840  test-rmse:1482.617432
## [107]    train-rmse:1475.370850  test-rmse:1482.606201
## [108]    train-rmse:1475.318237  test-rmse:1482.612183
## [109]    train-rmse:1475.211548  test-rmse:1482.651123
## [110]    train-rmse:1475.136841  test-rmse:1482.650513
## [111]    train-rmse:1475.044434  test-rmse:1482.683350
## [112]    train-rmse:1474.854370  test-rmse:1482.652222
## [113]    train-rmse:1474.845093  test-rmse:1482.666626
## [114]    train-rmse:1474.771851  test-rmse:1482.673340

## [115]    train-rmse:1474.710083  test-rmse:1482.696411
## [116]    train-rmse:1474.448853  test-rmse:1482.710327
## [117]    train-rmse:1474.409912  test-rmse:1482.703247
## [118]    train-rmse:1474.326416  test-rmse:1482.705933
## [119]    train-rmse:1474.317505  test-rmse:1482.710571
## [120]    train-rmse:1474.211304  test-rmse:1482.759766
## [121]    train-rmse:1474.160400  test-rmse:1482.756836
## [122]    train-rmse:1474.089844  test-rmse:1482.786255
## [123]    train-rmse:1474.056763  test-rmse:1482.800415
## [124]    train-rmse:1473.949097  test-rmse:1482.812622
## [125]    train-rmse:1473.857300  test-rmse:1482.822266
## Stopping. Best iteration:
## [95] train-rmse:1476.702759  test-rmse:1482.494141
```

Model 3

```
xgb_model3 <- xgb.train(data=dtrain,booster='gbtree',nrounds=1000,max_depth=8,eval_me
                eta=0.3,watchlist=w,early_stopping_rounds = 30)
```

```
## [1]  train-rmse:5440.340820  test-rmse:5432.918457
## Multiple eval metrics are present. Will use test_rmse for early stopping.
## Will train until test_rmse hasn't improved in 30 rounds.
##
## [2]  train-rmse:3954.926514  test-rmse:3948.443359
## [3]  train-rmse:2965.939453  test-rmse:2960.512451
## [4]  train-rmse:2331.679443  test-rmse:2327.384277
## [5]  train-rmse:1945.849121  test-rmse:1943.007935
## [6]  train-rmse:1724.538696  test-rmse:1723.770020
## [7]  train-rmse:1604.409790  test-rmse:1605.345825
## [8]  train-rmse:1541.515991  test-rmse:1544.165161
## [9]  train-rmse:1509.409180  test-rmse:1513.099609
## [10] train-rmse:1493.105591  test-rmse:1497.815186
## [11] train-rmse:1484.581665  test-rmse:1490.431030
## [12] train-rmse:1479.996948  test-rmse:1486.637207
## [13] train-rmse:1477.540527  test-rmse:1484.813843
## [14] train-rmse:1475.854614  test-rmse:1483.866455
## [15] train-rmse:1474.798462  test-rmse:1483.363281
## [16] train-rmse:1474.195435  test-rmse:1483.184937
## [17] train-rmse:1473.576050  test-rmse:1483.128174
## [18] train-rmse:1472.976807  test-rmse:1483.225098
## [19] train-rmse:1471.940063  test-rmse:1483.484375
## [20] train-rmse:1471.549561  test-rmse:1483.557739
## [21] train-rmse:1471.477295  test-rmse:1483.534180
## [22] train-rmse:1470.573242  test-rmse:1483.683105
## [23] train-rmse:1470.409668  test-rmse:1483.687744
## [24] train-rmse:1469.707642  test-rmse:1483.970703
## [25] train-rmse:1469.168091  test-rmse:1484.071899
## [26] train-rmse:1468.714111  test-rmse:1484.244751
## [27] train-rmse:1468.216064  test-rmse:1484.176147
## [28] train-rmse:1467.614624  test-rmse:1484.246582
## [29] train-rmse:1467.273926  test-rmse:1484.493408
## [30] train-rmse:1466.640015  test-rmse:1484.420410
## [31] train-rmse:1466.605713  test-rmse:1484.446899
## [32] train-rmse:1466.154175  test-rmse:1484.578247
## [33] train-rmse:1465.684937  test-rmse:1484.725708
## [34] train-rmse:1465.255493  test-rmse:1485.061279
## [35] train-rmse:1464.751831  test-rmse:1485.157715
## [36] train-rmse:1464.244629  test-rmse:1485.237427
## [37] train-rmse:1463.509277  test-rmse:1485.285522
## [38] train-rmse:1462.847900  test-rmse:1485.435059
```

```
## [39] train-rmse:1462.583008  test-rmse:1485.497681
## [40] train-rmse:1462.351318  test-rmse:1485.634644
## [41] train-rmse:1461.987549  test-rmse:1485.695068
## [42] train-rmse:1461.586670  test-rmse:1485.759766
## [43] train-rmse:1461.222290  test-rmse:1485.864868
## [44] train-rmse:1460.904541  test-rmse:1485.890503
## [45] train-rmse:1460.663452  test-rmse:1485.884399
## [46] train-rmse:1460.161255  test-rmse:1486.108398
## [47] train-rmse:1459.611450  test-rmse:1486.265869
## Stopping. Best iteration:
## [17] train-rmse:1473.576050  test-rmse:1483.128174
```

Model 4

```
xgb_model4 <- xgb.train(data=dtrain,booster='gbtree',nrounds=800,max_depth=6,eval_met
                        eta=0.135,watchlist=w,early_stopping_rounds = 30)
```

```
## [1]  train-rmse:6634.682129  test-rmse:6626.620117
## Multiple eval metrics are present. Will use test_rmse for early stopping.
## Will train until test_rmse hasn't improved in 30 rounds.
##
## [2]  train-rmse:5789.277344  test-rmse:5781.732910
## [3]  train-rmse:5064.992676  test-rmse:5057.577148
## [4]  train-rmse:4446.244141  test-rmse:4438.996094
## [5]  train-rmse:3919.754395  test-rmse:3912.774658
## [6]  train-rmse:3473.498047  test-rmse:3466.785645
## [7]  train-rmse:3097.755127  test-rmse:3091.212158
## [8]  train-rmse:2783.359863  test-rmse:2777.112061
## [9]  train-rmse:2522.162842  test-rmse:2516.188477
## [10] train-rmse:2307.276611  test-rmse:2301.458740
## [11] train-rmse:2132.010254  test-rmse:2126.546387
## [12] train-rmse:1990.651245  test-rmse:1985.732056
## [13] train-rmse:1877.574097  test-rmse:1873.045654
## [14] train-rmse:1788.107666  test-rmse:1784.157349
## [15] train-rmse:1718.046509  test-rmse:1714.492432
## [16] train-rmse:1663.198120  test-rmse:1660.026611
## [17] train-rmse:1621.156860  test-rmse:1618.375000
## [18] train-rmse:1588.680908  test-rmse:1586.194214
## [19] train-rmse:1563.844482  test-rmse:1561.676758
## [20] train-rmse:1544.490356  test-rmse:1542.672729
## [21] train-rmse:1529.903809  test-rmse:1528.286987
## [22] train-rmse:1518.926392  test-rmse:1517.564819
## [23] train-rmse:1510.503662  test-rmse:1509.443237
## [24] train-rmse:1504.017578  test-rmse:1503.202881
## [25] train-rmse:1499.115356  test-rmse:1498.530884
```

```
## [26] train-rmse:1495.409546  test-rmse:1495.086426
## [27] train-rmse:1492.302734  test-rmse:1492.264038
## [28] train-rmse:1490.074829  test-rmse:1490.200684
## [29] train-rmse:1488.334839  test-rmse:1488.603516
## [30] train-rmse:1487.016357  test-rmse:1487.410889
## [31] train-rmse:1486.035156  test-rmse:1486.546143
## [32] train-rmse:1485.204956  test-rmse:1485.897827
## [33] train-rmse:1484.603760  test-rmse:1485.466064
## [34] train-rmse:1483.944946  test-rmse:1484.964355
## [35] train-rmse:1483.528198  test-rmse:1484.614380
## [36] train-rmse:1483.107300  test-rmse:1484.356079
## [37] train-rmse:1482.877441  test-rmse:1484.208374
## [38] train-rmse:1482.321411  test-rmse:1483.835205
## [39] train-rmse:1481.987549  test-rmse:1483.736572
## [40] train-rmse:1481.773315  test-rmse:1483.578979
## [41] train-rmse:1481.643311  test-rmse:1483.508545
## [42] train-rmse:1481.528564  test-rmse:1483.456177
## [43] train-rmse:1481.399658  test-rmse:1483.367432
## [44] train-rmse:1481.279297  test-rmse:1483.278198
## [45] train-rmse:1480.977417  test-rmse:1483.154785
## [46] train-rmse:1480.435791  test-rmse:1482.873779
## [47] train-rmse:1480.368530  test-rmse:1482.853027
## [48] train-rmse:1480.276978  test-rmse:1482.828491
## [49] train-rmse:1480.218262  test-rmse:1482.830200
## [50] train-rmse:1480.175537  test-rmse:1482.816162
## [51] train-rmse:1479.798096  test-rmse:1482.705078
## [52] train-rmse:1479.721802  test-rmse:1482.669922
## [53] train-rmse:1479.648560  test-rmse:1482.660400
## [54] train-rmse:1479.234375  test-rmse:1482.427612
## [55] train-rmse:1479.218750  test-rmse:1482.422119

## [56] train-rmse:1479.158936  test-rmse:1482.390137
## [57] train-rmse:1479.041626  test-rmse:1482.431885
## [58] train-rmse:1478.637695  test-rmse:1482.372314
## [59] train-rmse:1478.586792  test-rmse:1482.375122
## [60] train-rmse:1478.544678  test-rmse:1482.349609
## [61] train-rmse:1478.389526  test-rmse:1482.393311
## [62] train-rmse:1478.271484  test-rmse:1482.411255
## [63] train-rmse:1478.246704  test-rmse:1482.417480
## [64] train-rmse:1477.953003  test-rmse:1482.489990
## [65] train-rmse:1477.684204  test-rmse:1482.484985
## [66] train-rmse:1477.380249  test-rmse:1482.369995
## [67] train-rmse:1477.175049  test-rmse:1482.414551
## [68] train-rmse:1477.139038  test-rmse:1482.434937
## [69] train-rmse:1476.926147  test-rmse:1482.440674
## [70] train-rmse:1476.747192  test-rmse:1482.447632
## [71] train-rmse:1476.613403  test-rmse:1482.486450
## [72] train-rmse:1476.564209  test-rmse:1482.488403
## [73] train-rmse:1476.432007  test-rmse:1482.480469
## [74] train-rmse:1476.408691  test-rmse:1482.501953
## [75] train-rmse:1476.275269  test-rmse:1482.504517
## [76] train-rmse:1476.157227  test-rmse:1482.471924
```

```
## [77] train-rmse:1475.915771  test-rmse:1482.496948
## [78] train-rmse:1475.853882  test-rmse:1482.509888
## [79] train-rmse:1475.682007  test-rmse:1482.569092
## [80] train-rmse:1475.516602  test-rmse:1482.543579
## [81] train-rmse:1475.419434  test-rmse:1482.613281
## [82] train-rmse:1475.297119  test-rmse:1482.603638
## [83] train-rmse:1475.156860  test-rmse:1482.669556
## [84] train-rmse:1475.048706  test-rmse:1482.717407
## [85] train-rmse:1474.936646  test-rmse:1482.704346
## [86] train-rmse:1474.883301  test-rmse:1482.735718
## [87] train-rmse:1474.783203  test-rmse:1482.764771
## [88] train-rmse:1474.639282  test-rmse:1482.803467
## [89] train-rmse:1474.422485  test-rmse:1482.839478
## [90] train-rmse:1474.345337  test-rmse:1482.854126
## Stopping. Best iteration:
## [60] train-rmse:1478.544678  test-rmse:1482.349609
```

Model 2 is the best model

```
best_model <- xgb.train(data=dtrain,booster='gbtree',nrounds=95,max_depth=6,eval_metr
                        eta=0.1,watchlist=w)
```

```
## [1]  train-rmse:6889.255859  test-rmse:6881.096191
## [2]  train-rmse:6235.568359  test-rmse:6227.811035
## [3]  train-rmse:5650.816406  test-rmse:5643.126465
## [4]  train-rmse:5128.273438  test-rmse:5120.769043
## [5]  train-rmse:4662.144531  test-rmse:4654.789062
## [6]  train-rmse:4247.190918  test-rmse:4240.054688
## [7]  train-rmse:3878.384277  test-rmse:3871.440918
## [8]  train-rmse:3551.668457  test-rmse:3544.884521
## [9]  train-rmse:3263.005371  test-rmse:3256.394775
## [10] train-rmse:3008.710205  test-rmse:3002.257568
## [11] train-rmse:2785.531982  test-rmse:2779.270996
## [12] train-rmse:2590.678955  test-rmse:2584.555176
## [13] train-rmse:2420.921631  test-rmse:2415.067139
## [14] train-rmse:2274.069580  test-rmse:2268.491211
## [15] train-rmse:2147.823242  test-rmse:2142.476074
## [16] train-rmse:2039.728638  test-rmse:2034.714600
## [17] train-rmse:1947.488525  test-rmse:1942.823608
## [18] train-rmse:1869.466187  test-rmse:1865.267334
## [19] train-rmse:1803.599365  test-rmse:1799.663330
## [20] train-rmse:1748.241577  test-rmse:1744.615234
## [21] train-rmse:1702.010376  test-rmse:1698.643921
## [22] train-rmse:1663.474854  test-rmse:1660.449219
```

```
## [23] train-rmse:1631.571655  test-rmse:1628.812622
## [24] train-rmse:1605.214355  test-rmse:1602.724854
## [25] train-rmse:1583.494019  test-rmse:1581.223389
## [26] train-rmse:1565.523071  test-rmse:1563.539062
## [27] train-rmse:1550.761719  test-rmse:1548.969360
## [28] train-rmse:1538.708740  test-rmse:1537.119141
## [29] train-rmse:1528.764893  test-rmse:1527.373901
## [30] train-rmse:1520.502686  test-rmse:1519.357300
## [31] train-rmse:1513.879272  test-rmse:1512.943726
## [32] train-rmse:1508.387695  test-rmse:1507.603394
## [33] train-rmse:1503.703003  test-rmse:1503.106323
## [34] train-rmse:1500.057251  test-rmse:1499.618652
## [35] train-rmse:1497.002808  test-rmse:1496.733643
## [36] train-rmse:1494.351807  test-rmse:1494.251587
## [37] train-rmse:1492.260132  test-rmse:1492.264771
## [38] train-rmse:1490.525513  test-rmse:1490.682617
## [39] train-rmse:1489.122437  test-rmse:1489.425659
## [40] train-rmse:1487.895752  test-rmse:1488.400024
## [41] train-rmse:1486.893555  test-rmse:1487.472656
## [42] train-rmse:1486.109375  test-rmse:1486.753296
## [43] train-rmse:1485.450806  test-rmse:1486.186035
## [44] train-rmse:1484.875122  test-rmse:1485.699219
## [45] train-rmse:1484.413574  test-rmse:1485.268677
## [46] train-rmse:1484.012939  test-rmse:1484.894287
## [47] train-rmse:1483.663940  test-rmse:1484.628906
## [48] train-rmse:1483.265015  test-rmse:1484.412964
## [49] train-rmse:1483.013428  test-rmse:1484.256958
## [50] train-rmse:1482.714966  test-rmse:1483.999268
## [51] train-rmse:1482.488037  test-rmse:1483.877930
## [52] train-rmse:1482.296143  test-rmse:1483.766113
## [53] train-rmse:1482.082642  test-rmse:1483.629028
## [54] train-rmse:1481.976196  test-rmse:1483.578979
## [55] train-rmse:1481.611206  test-rmse:1483.296143
## [56] train-rmse:1481.505127  test-rmse:1483.250854
## [57] train-rmse:1481.403931  test-rmse:1483.222900
## [58] train-rmse:1481.283936  test-rmse:1483.166870
## [59] train-rmse:1481.212891  test-rmse:1483.149414
## [60] train-rmse:1481.070190  test-rmse:1483.070923
## [61] train-rmse:1480.937256  test-rmse:1483.070923
## [62] train-rmse:1480.651245  test-rmse:1483.030029
## [63] train-rmse:1480.608521  test-rmse:1483.012451
## [64] train-rmse:1480.321045  test-rmse:1482.943481
## [65] train-rmse:1480.252441  test-rmse:1482.939697
## [66] train-rmse:1480.141235  test-rmse:1482.911621
## [67] train-rmse:1480.067017  test-rmse:1482.885620
## [68] train-rmse:1479.807373  test-rmse:1482.826904
## [69] train-rmse:1479.748169  test-rmse:1482.829346
## [70] train-rmse:1479.675781  test-rmse:1482.828857
## [71] train-rmse:1479.421509  test-rmse:1482.688354
## [72] train-rmse:1479.145752  test-rmse:1482.586304
## [73] train-rmse:1479.020264  test-rmse:1482.552002
```

```
## [74] train-rmse:1478.737427  test-rmse:1482.543823
## [75] train-rmse:1478.701538  test-rmse:1482.539429
## [76] train-rmse:1478.588379  test-rmse:1482.532837
## [77] train-rmse:1478.470337  test-rmse:1482.496582
## [78] train-rmse:1478.454834  test-rmse:1482.502686
## [79] train-rmse:1478.345825  test-rmse:1482.500732
## [80] train-rmse:1478.194214  test-rmse:1482.523315
## [81] train-rmse:1478.089966  test-rmse:1482.512207
## [82] train-rmse:1477.859131  test-rmse:1482.498779
## [83] train-rmse:1477.651001  test-rmse:1482.531250
## [84] train-rmse:1477.638184  test-rmse:1482.538330
## [85] train-rmse:1477.595947  test-rmse:1482.551392
## [86] train-rmse:1477.534302  test-rmse:1482.561890
## [87] train-rmse:1477.422852  test-rmse:1482.570679
## [88] train-rmse:1477.266602  test-rmse:1482.534546
## [89] train-rmse:1477.175903  test-rmse:1482.536255
## [90] train-rmse:1477.134644  test-rmse:1482.532104
## [91] train-rmse:1477.068115  test-rmse:1482.510986
## [92] train-rmse:1477.004028  test-rmse:1482.513916
## [93] train-rmse:1476.895386  test-rmse:1482.497681
## [94] train-rmse:1476.753174  test-rmse:1482.496094
## [95] train-rmse:1476.702759  test-rmse:1482.494141
```

Prediction for test set

```
pred_sales <- predict(best_model,newdata = dtest,class='response')
pred_sales <- round(pred_sales)
head(pred_sales)
```

```
## [1]  6432  8728 12716  7291 10837  6287
```