IE5374Project

Dhaval Jariwala

2022-12-06

IE5374 PROJECT

BACKORDER PREDICTION

GROUP MEMBERS - DHAVAL JARIWALA, DISHANT GUPTA, YESHA GOSALIYA

CHECK THE DATASETS

```
#TRAINING DATA

head(trainData)
```

```
##   national_inv lead_time in_transit_qty forecast_3_month forecast_6_month
## 1            0        NA              0                0                0
## 2            2         9              0                0                0
## 3            2        NA              0                0                0
## 4            7         8              0                0                0
## 5            8        NA              0                0                0
## 6           13         8              0                0                0
##   forecast_9_month sales_1_month sales_3_month sales_6_month sales_9_month
## 1                0             0             0             0             0
## 2                0             0             0             0             0
## 3                0             0             0             0             0
## 4                0             0             0             0             0
## 5                0             0             0             0             4
## 6                0             0             0             0             0
##   min_bank potential_issue pieces_past_due perf_6_month_avg perf_12_month_avg
## 1        0              No               0           -99.00            -99.00
## 2        0              No               0             0.99              0.99
## 3        0              No               0           -99.00            -99.00
## 4        1              No               0             0.10              0.13
## 5        2              No               0           -99.00            -99.00
## 6        0              No               0             0.82              0.87
##   local_bo_qty deck_risk oe_constraint ppap_risk stop_auto_buy rev_stop
## 1            0        No            No        No           Yes       No
## 2            0        No            No        No           Yes       No
## 3            0       Yes            No        No           Yes       No
## 4            0        No            No        No           Yes       No
## 5            0       Yes            No        No           Yes       No
## 6            0        No            No        No           Yes       No
##   went_on_backorder
## 1                No
## 2                No
## 3                No
## 4                No
## 5                No
## 6                No
```

```
# Check dimensions of the dataset
dim(trainData)
```

```
## [1] 1687861      22
```

```
# check the datatype of each column
sapply(trainData,class)
```

```
##      national_inv         lead_time    in_transit_qty  forecast_3_month
##         "integer"         "integer"         "integer"         "integer"
##  forecast_6_month  forecast_9_month     sales_1_month     sales_3_month
##         "integer"         "integer"         "integer"         "integer"
##     sales_6_month     sales_9_month          min_bank   potential_issue
##         "integer"         "integer"         "integer"          "factor"
##   pieces_past_due  perf_6_month_avg perf_12_month_avg      local_bo_qty
##         "integer"         "numeric"         "numeric"         "integer"
##         deck_risk     oe_constraint         ppap_risk     stop_auto_buy
##          "factor"          "factor"          "factor"          "factor"
##          rev_stop went_on_backorder
##          "factor"          "factor"
```

```
# summary of all columns in the dataset
summary(trainData)
```

```
##   national_inv        lead_time      in_transit_qty     forecast_3_month
##  Min.   :  -27256   Min.   : 0.00    Min.   :     0.0   Min.   :      0.0
##  1st Qu.:       4   1st Qu.: 4.00    1st Qu.:     0.0   1st Qu.:      0.0
##  Median :      15   Median : 8.00    Median :     0.0   Median :      0.0
##  Mean   :     496   Mean   : 7.87    Mean   :    44.1   Mean   :    178.1
##  3rd Qu.:      80   3rd Qu.: 9.00    3rd Qu.:     0.0   3rd Qu.:      4.0
##  Max.   :12334404   Max.   :52.00    Max.   :489408.0   Max.   :1427612.0
##  NA's   :1          NA's   :100894   NA's   :1          NA's   :1
##  forecast_6_month  forecast_9_month  sales_1_month      sales_3_month
##  Min.   :      0   Min.   :      0   Min.   :     0.0   Min.   :      0
##  1st Qu.:      0   1st Qu.:      0   1st Qu.:     0.0   1st Qu.:      0
##  Median :      0   Median :      0   Median :     0.0   Median :      1
##  Mean   :    345   Mean   :    506   Mean   :    55.9   Mean   :    175
##  3rd Qu.:     12   3rd Qu.:     20   3rd Qu.:     4.0   3rd Qu.:     15
##  Max.   :2461360   Max.   :3777304   Max.   :741774.0   Max.   :1105478
##  NA's   :1         NA's   :1         NA's   :1          NA's   :1
##  sales_6_month       sales_9_month        min_bank         potential_issue
##  Min.   :      0.0   Min.   :      0   Min.   :     0.00      :      1
##  1st Qu.:      0.0   1st Qu.:      0   1st Qu.:     0.00   No :1686953
##  Median :      2.0   Median :      4   Median :     0.00   Yes:    907
##  Mean   :    341.7   Mean   :    525   Mean   :    52.77
##  3rd Qu.:     31.0   3rd Qu.:     47   3rd Qu.:     3.00
##  Max.   :2146625.0   Max.   :3205172   Max.   :313319.00
##  NA's   :1           NA's   :1         NA's   :1
##  pieces_past_due     perf_6_month_avg  perf_12_month_avg  local_bo_qty
##  Min.   :     0.00   Min.   :-99.000   Min.   :-99.000   Min.   :    0.000
##  1st Qu.:     0.00   1st Qu.:  0.630   1st Qu.:  0.660   1st Qu.:    0.000
##  Median :     0.00   Median :  0.820   Median :  0.810   Median :    0.000
##  Mean   :     2.04   Mean   : -6.872   Mean   : -6.438   Mean   :    0.626
##  3rd Qu.:     0.00   3rd Qu.:  0.970   3rd Qu.:  0.950   3rd Qu.:    0.000
##  Max.   :146496.00   Max.   :  1.000   Max.   :  1.000   Max.   :12530.000
##  NA's   :1           NA's   :1         NA's   :1         NA's   :1
##  deck_risk     oe_constraint ppap_risk     stop_auto_buy rev_stop
##     :      1      :      1      :      1      :      1      :      1
##  No :1300377   No :1687615   No :1484026   No :  61086   No :1687129
##  Yes: 387483   Yes:    245   Yes: 203834   Yes:1626774   Yes:    731
##
##
##
##
##  went_on_backorder
##     :      1
##  No :1676567
##  Yes:  11293
##
##
##
##
```

```
# TESTING DATA

head(testData)
```

```
##   national_inv lead_time in_transit_qty forecast_3_month forecast_6_month
## 1           62        NA              0                0                0
## 2            9        NA              0                0                0
## 3           17         8              0                0                0
## 4            9         2              0                0                0
## 5            2         8              0                0                0
## 6           15         2              0                0                0
##   forecast_9_month sales_1_month sales_3_month sales_6_month sales_9_month
## 1                0             0             0             0             0
## 2                0             0             0             0             0
## 3                0             0             0             0             0
## 4                0             0             0             0             2
## 5                0             0             0             0             0
## 6                0             0             0             1             2
##   min_bank potential_issue pieces_past_due perf_6_month_avg perf_12_month_avg
## 1        1              No               0           -99.00            -99.00
## 2        1              No               0           -99.00            -99.00
## 3        0              No               0             0.92              0.95
## 4        0              No               0             0.78              0.75
## 5        0              No               0             0.54              0.71
## 6        0              No               0             0.37              0.68
##   local_bo_qty deck_risk oe_constraint ppap_risk stop_auto_buy rev_stop
## 1            0       Yes            No        No           Yes       No
## 2            0        No            No       Yes            No       No
## 3            0        No            No        No           Yes       No
## 4            0        No            No       Yes           Yes       No
## 5            0        No            No        No           Yes       No
## 6            0        No            No        No           Yes       No
##   went_on_backorder
## 1                No
## 2                No
## 3                No
## 4                No
## 5                No
## 6                No
```

```
# Check dimensions of the dataset
dim(testData)
```

```
## [1] 242076     22
```

```
# check the datatype of each column
sapply(testData,class)
```

```
##      national_inv         lead_time    in_transit_qty  forecast_3_month
##         "integer"         "integer"         "integer"         "integer"
##  forecast_6_month  forecast_9_month     sales_1_month     sales_3_month
##         "integer"         "integer"         "integer"         "integer"
##     sales_6_month     sales_9_month          min_bank   potential_issue
##         "integer"         "integer"         "integer"          "factor"
##   pieces_past_due  perf_6_month_avg perf_12_month_avg      local_bo_qty
##         "integer"         "numeric"         "numeric"         "integer"
##         deck_risk     oe_constraint         ppap_risk     stop_auto_buy
##          "factor"          "factor"          "factor"          "factor"
##          rev_stop went_on_backorder
##          "factor"          "factor"
```

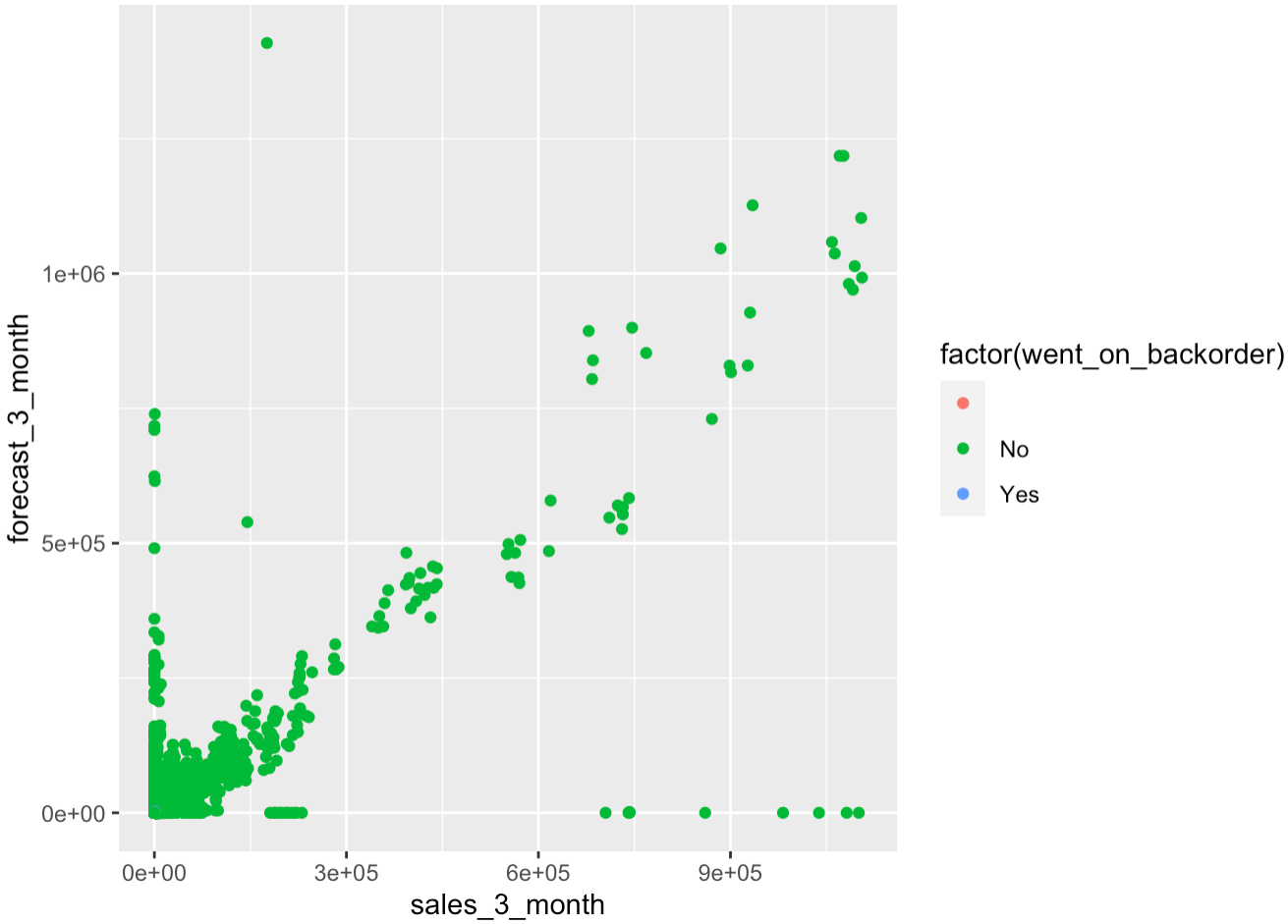
```
# summary of all columns in the dataset
summary(testData)
```

```
##   national_inv        lead_time      in_transit_qty      forecast_3_month
##  Min.   :  -25414   Min.   : 0.000   Min.   :     0.00   Min.   :      0.0
##  1st Qu.:       4   1st Qu.: 4.000   1st Qu.:     0.00   1st Qu.:      0.0
##  Median :      15   Median : 8.000   Median :     0.00   Median :      0.0
##  Mean   :     500   Mean   : 7.923   Mean   :    36.18   Mean   :    181.5
##  3rd Qu.:      81   3rd Qu.: 9.000   3rd Qu.:     0.00   3rd Qu.:      4.0
##  Max.   :12145792   Max.   :52.000   Max.   :265272.00   Max.   :1510592.0
##  NA's   :1          NA's   :14725    NA's   :1           NA's   :1
##  forecast_6_month    forecast_9_month    sales_1_month      sales_3_month
##  Min.   :      0.0   Min.   :      0.0   Min.   :     0.0   Min.   :      0.0
##  1st Qu.:      0.0   1st Qu.:      0.0   1st Qu.:     0.0   1st Qu.:      0.0
##  Median :      0.0   Median :      0.0   Median :     0.0   Median :      1.0
##  Mean   :    348.8   Mean   :    508.3   Mean   :    51.5   Mean   :    172.1
##  3rd Qu.:     12.0   3rd Qu.:     20.0   3rd Qu.:     4.0   3rd Qu.:     14.0
##  Max.   :2157024.0   Max.   :3162260.0   Max.   :349620.0   Max.   :1099852.0
##  NA's   :1           NA's   :1           NA's   :1          NA's   :1
##  sales_6_month       sales_9_month        min_bank        potential_issue
##  Min.   :      0.0   Min.   :      0   Min.   :     0.0      :     1
##  1st Qu.:      0.0   1st Qu.:      0   1st Qu.:     0.0   No :241993
##  Median :      2.0   Median :      4   Median :     0.0   Yes:    82
##  Mean   :    340.4   Mean   :    512   Mean   :    52.8
##  3rd Qu.:     30.0   3rd Qu.:     46   3rd Qu.:     3.0
##  Max.   :2103389.0   Max.   :3195211   Max.   :303713.0
##  NA's   :1           NA's   :1         NA's   :1
##  pieces_past_due    perf_6_month_avg  perf_12_month_avg  local_bo_qty
##  Min.   :    0.00   Min.   :-99.000   Min.   :-99.000   Min.   :   0.000
##  1st Qu.:    0.00   1st Qu.:  0.630   1st Qu.:  0.660   1st Qu.:   0.000
##  Median :    0.00   Median :  0.820   Median :  0.810   Median :   0.000
##  Mean   :    1.82   Mean   : -7.094   Mean   : -6.632   Mean   :   0.844
##  3rd Qu.:    0.00   3rd Qu.:  0.960   3rd Qu.:  0.950   3rd Qu.:   0.000
##  Max.   :79964.00   Max.   :  1.000   Max.   :  1.000   Max.   :6232.000
##  NA's   :1          NA's   :1         NA's   :1         NA's   :1
##  deck_risk    oe_constraint ppap_risk    stop_auto_buy rev_stop
##     :     1      :     1       :     1      :     1       :     1
##  No :194105   No :242028    No :213357   No :  9458    No :241967
##  Yes: 47970   Yes:    47    Yes: 28718   Yes:232617    Yes:   108
##
##
##
##
##  went_on_backorder
##     :     1
##  No :239387
##  Yes:  2688
##
##
##
##
```

EDA

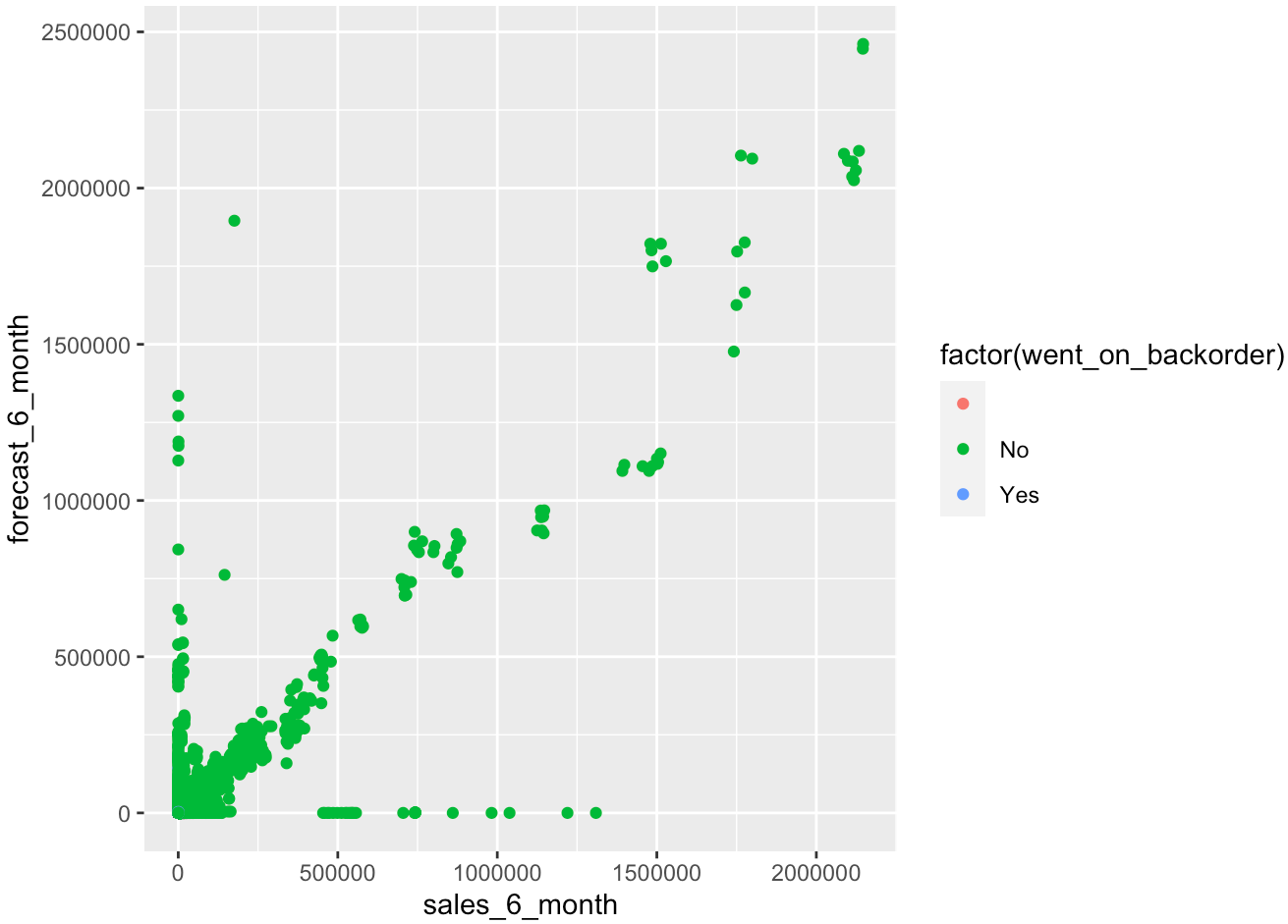
```
ggplot(data = trainData, aes(x = sales_3_month, y = forecast_3_month)) + geom_point(a
es(color = factor(went_on_backorder)))
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



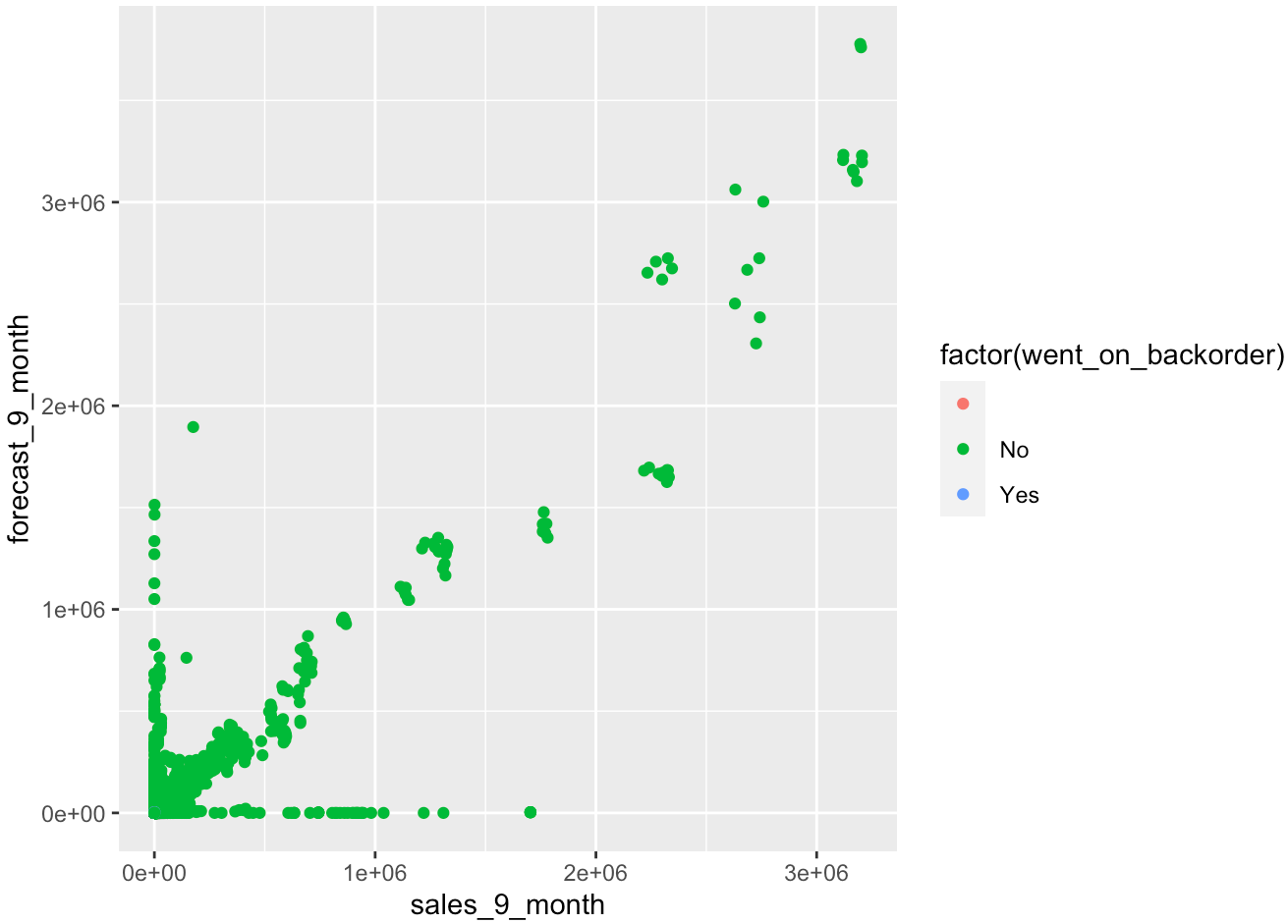

```
ggplot(data = trainData, aes(x = sales_6_month, y = forecast_6_month)) + geom_point(a
es(color = factor(went_on_backorder)))
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



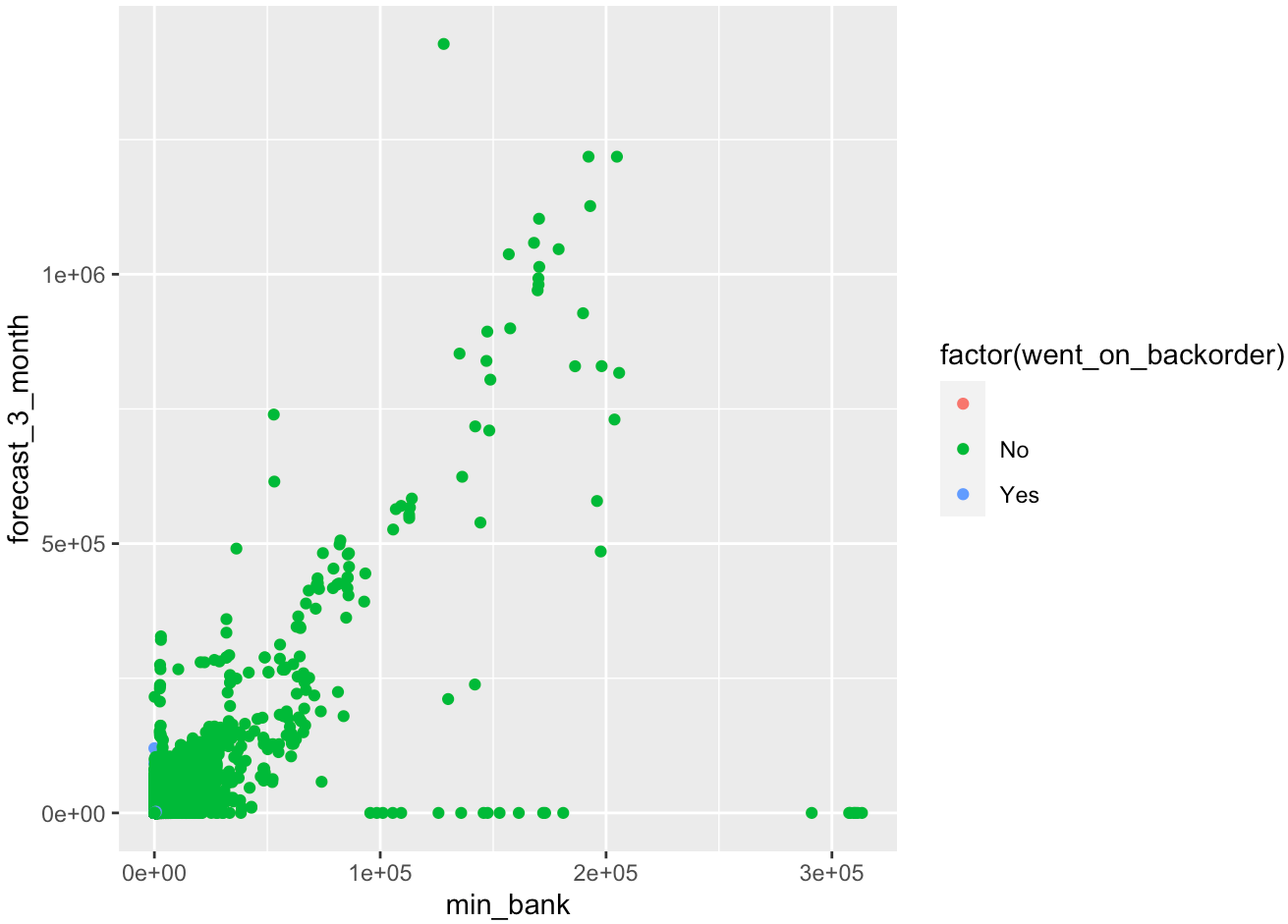

```
ggplot(data = trainData ,aes(x = sales_9_month, y = forecast_9_month)) + geom_point(a
es(color = factor(went_on_backorder)))
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



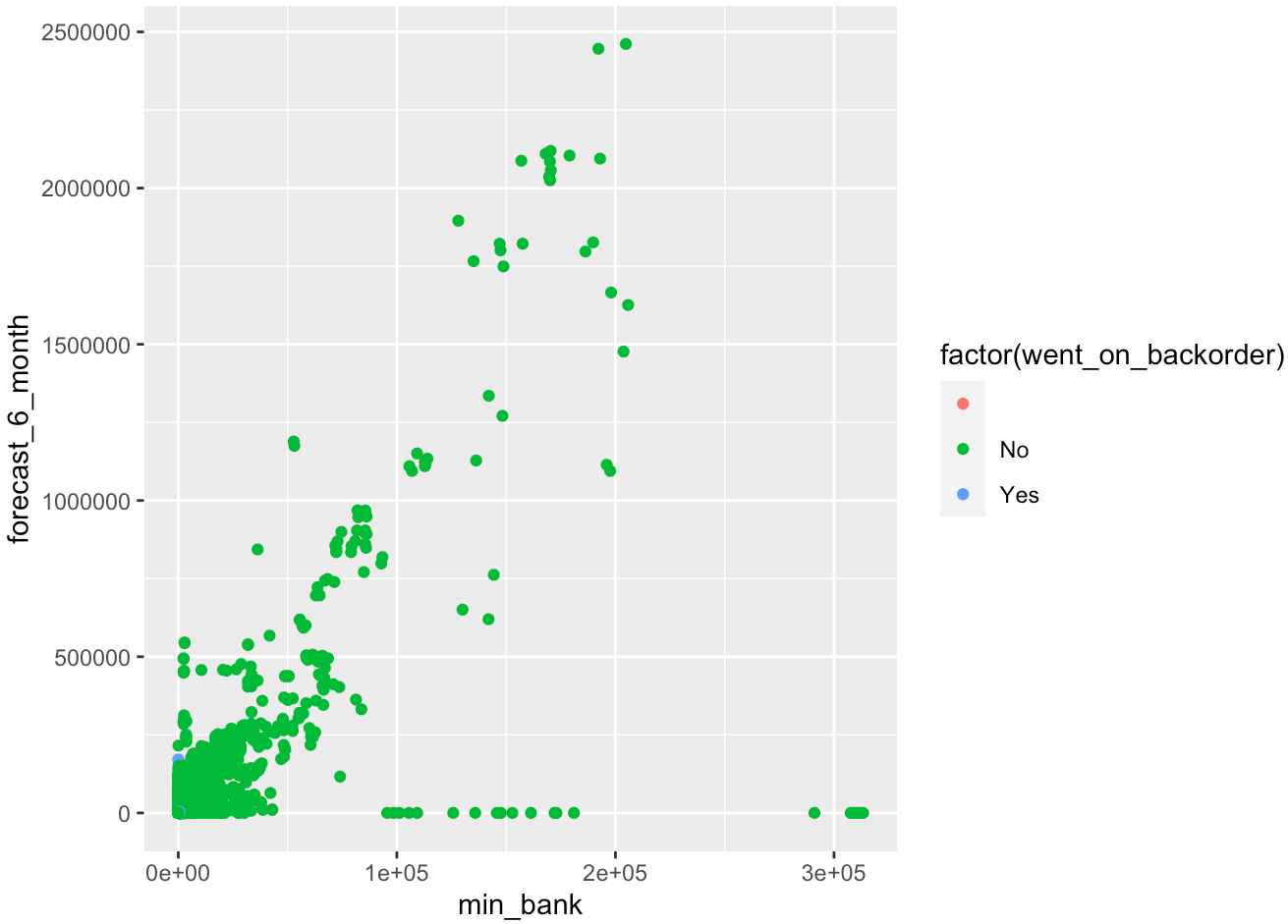

```
ggplot(data = trainData, aes(x = min_bank, y = forecast_3_month)) + geom_point(aes(co
lor = factor(went_on_backorder)))
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



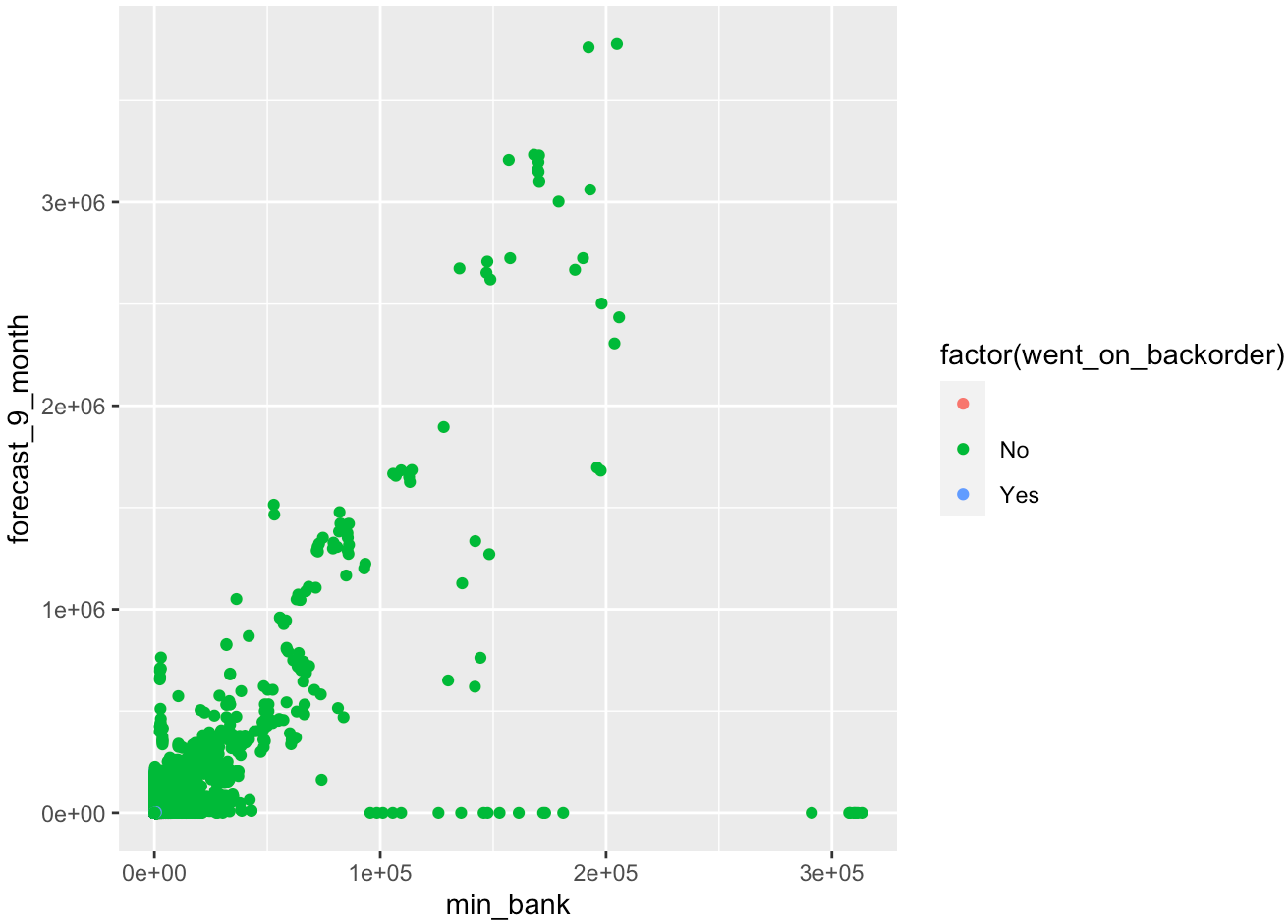

```
ggplot(data = trainData, aes(x = min_bank, y = forecast_6_month)) + geom_point(aes(co
lor = factor(went_on_backorder)))
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



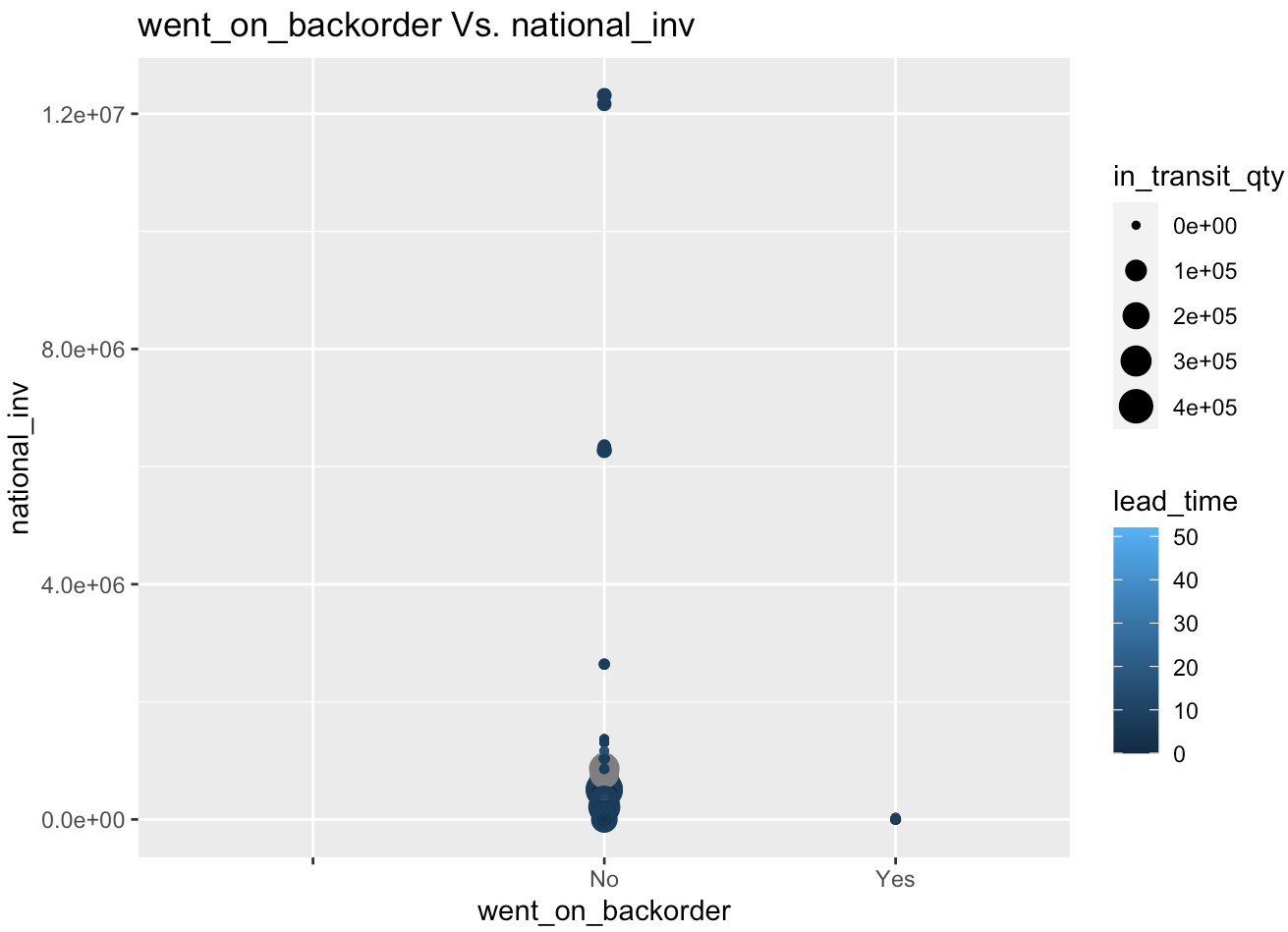

```
ggplot(data = trainData, aes(x = min_bank, y = forecast_9_month)) + geom_point(aes(co
lor = factor(went_on_backorder)))
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```




```
ggplot(data=trainData, aes(x=went_on_backorder, y=national_inv, colour = lead_time, s
ize = in_transit_qty)) +
    geom_point() +
    xlab("went_on_backorder")+ylab("national_inv") +
    ggtitle("went_on_backorder Vs. national_inv")
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



DATA CLEANING

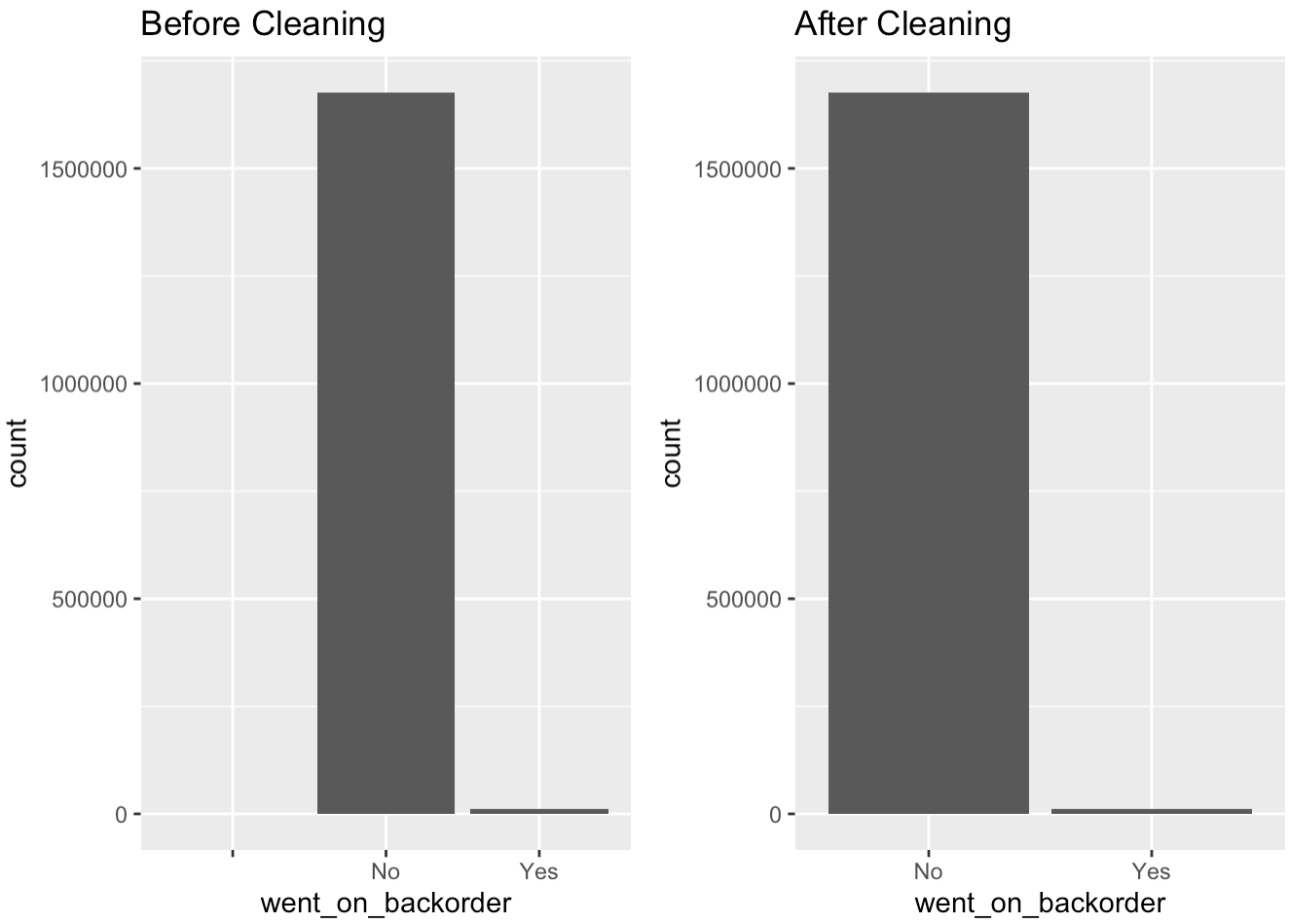
```
# TRAINING DATA

#checking the outcome variable for mislabled entries
rawTrainData = trainData
p1 = ggplot(rawTrainData, aes(went_on_backorder)) + geom_bar() + ggtitle("Before Clea
ning")

# from the summary it is evident that there is one row which is incorrectly entered i
n the dataset, it has either missing or null values
trainData <- trainData %>% filter(went_on_backorder != "")

# remove the rows with all NA
trainData <- trainData[apply(trainData, 1, function(y) !all(is.na(y))),]

# remove the Cols with all NA
trainData <- trainData[sapply(trainData, function(x) !all(is.na(x)))]

# drop unused levels induced by null row
trainData <- droplevels(trainData)

dim(trainData)
```

```
## [1] 1687860      22
```

```
summary(trainData)
```

```
##   national_inv        lead_time      in_transit_qty     forecast_3_month
##  Min.   :  -27256   Min.   : 0.00    Min.   :     0.0   Min.   :      0.0
##  1st Qu.:       4   1st Qu.: 4.00    1st Qu.:     0.0   1st Qu.:      0.0
##  Median :      15   Median : 8.00    Median :     0.0   Median :      0.0
##  Mean   :     496   Mean   : 7.87    Mean   :    44.1   Mean   :    178.1
##  3rd Qu.:      80   3rd Qu.: 9.00    3rd Qu.:     0.0   3rd Qu.:      4.0
##  Max.   :12334404   Max.   :52.00    Max.   :489408.0   Max.   :1427612.0
##                     NA's   :100893
##  forecast_6_month  forecast_9_month  sales_1_month      sales_3_month
##  Min.   :      0   Min.   :      0   Min.   :     0.0   Min.   :      0
##  1st Qu.:      0   1st Qu.:      0   1st Qu.:     0.0   1st Qu.:      0
##  Median :      0   Median :      0   Median :     0.0   Median :      1
##  Mean   :    345   Mean   :    506   Mean   :    55.9   Mean   :    175
##  3rd Qu.:     12   3rd Qu.:     20   3rd Qu.:     4.0   3rd Qu.:     15
##  Max.   :2461360   Max.   :3777304   Max.   :741774.0   Max.   :1105478
##
##  sales_6_month       sales_9_month        min_bank         potential_issue
##  Min.   :      0.0   Min.   :      0   Min.   :     0.00   No :1686953
##  1st Qu.:      0.0   1st Qu.:      0   1st Qu.:     0.00   Yes:    907
##  Median :      2.0   Median :      4   Median :     0.00
##  Mean   :    341.7   Mean   :    525   Mean   :    52.77
##  3rd Qu.:     31.0   3rd Qu.:     47   3rd Qu.:     3.00
##  Max.   :2146625.0   Max.   :3205172   Max.   :313319.00
##
##  pieces_past_due     perf_6_month_avg  perf_12_month_avg  local_bo_qty
##  Min.   :     0.00   Min.   :-99.000   Min.   :-99.000   Min.   :    0.000
##  1st Qu.:     0.00   1st Qu.:  0.630   1st Qu.:  0.660   1st Qu.:    0.000
##  Median :     0.00   Median :  0.820   Median :  0.810   Median :    0.000
##  Mean   :     2.04   Mean   : -6.872   Mean   : -6.438   Mean   :    0.626
##  3rd Qu.:     0.00   3rd Qu.:  0.970   3rd Qu.:  0.950   3rd Qu.:    0.000
##  Max.   :146496.00   Max.   :  1.000   Max.   :  1.000   Max.   :12530.000
##
##  deck_risk     oe_constraint ppap_risk     stop_auto_buy rev_stop
##  No :1300377   No :1687615   No :1484026   No :  61086   No :1687129
##  Yes: 387483   Yes:    245   Yes: 203834   Yes:1626774   Yes:    731
##
##
##
##
##
##  went_on_backorder
##  No :1676567
##  Yes:  11293
##
##
##
##
##
```

```
p2 = ggplot(trainData, aes(went_on_backorder)) + geom_bar() + ggtitle("After Cleanin
g")
grid.arrange(p1, p2, ncol = 2)
```




```
# rows and cols with all NA values and un-used levels are removed from the data set
```

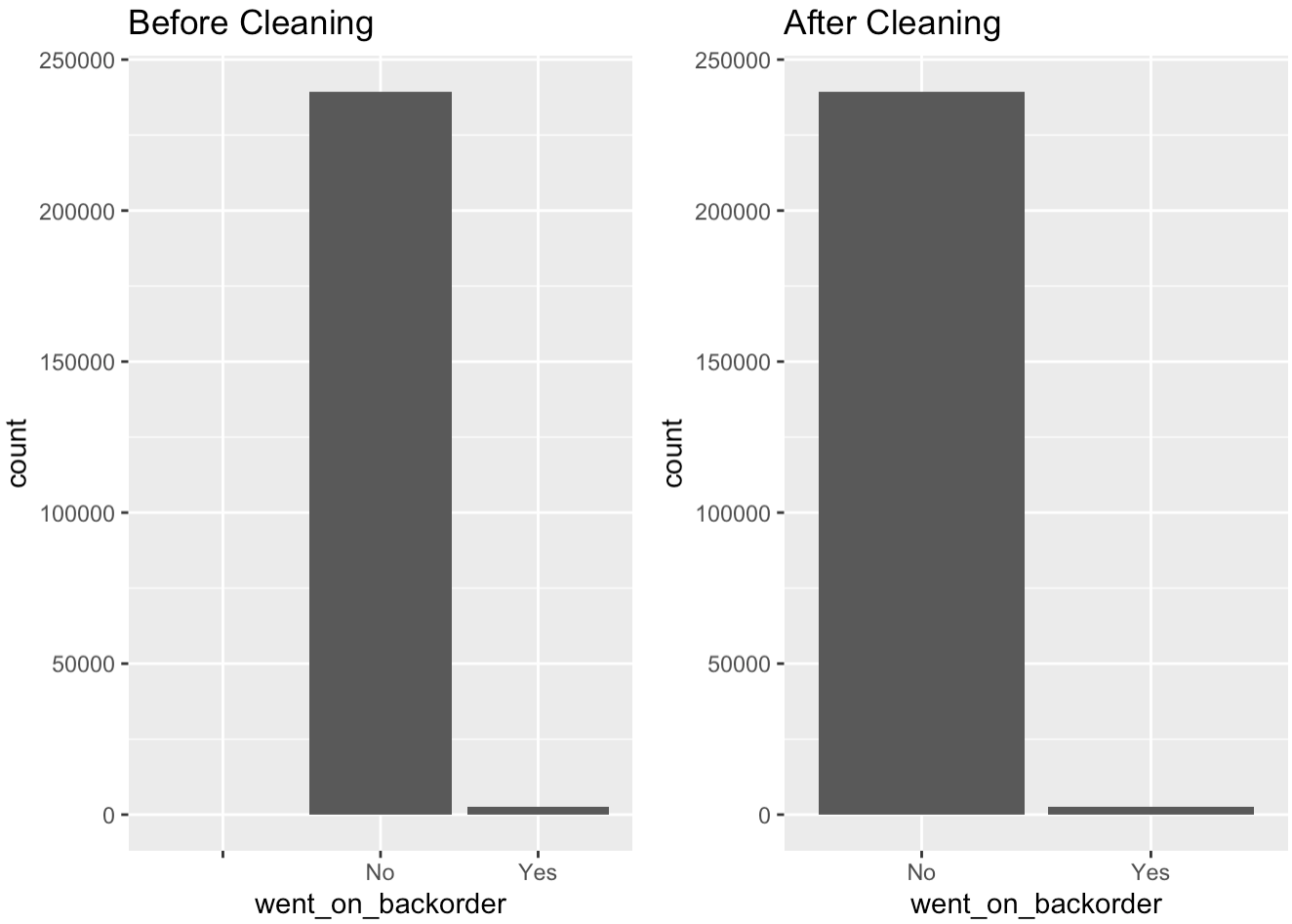
```
# TESTING DATA

#checking the outcome variable for mislabled entries
rawTestData = testData
p1 = ggplot(rawTestData, aes(went_on_backorder)) + geom_bar() + ggtitle("Before Clean
ing")

# from the summary it is evident that there is one row which is incorrectly entered i
n the dataset, it has either missing or null values
testData <- testData %>% filter(went_on_backorder != "")

# remove the rows with all NA
testData <- testData[apply(testData, 1, function(y) !all(is.na(y))),]

# remove the Cols with all NA
testData <- testData[sapply(testData, function(x) !all(is.na(x)))]

# drop unused levels induced by null row
testData <- droplevels(testData)

dim(testData)
```

```
## [1] 242075     22
```

```
summary(testData)
```

```
##   national_inv        lead_time      in_transit_qty      forecast_3_month
##  Min.   :  -25414   Min.   : 0.000   Min.   :     0.00   Min.   :      0.0
##  1st Qu.:       4   1st Qu.: 4.000   1st Qu.:     0.00   1st Qu.:      0.0
##  Median :      15   Median : 8.000   Median :     0.00   Median :      0.0
##  Mean   :     500   Mean   : 7.923   Mean   :    36.18   Mean   :    181.5
##  3rd Qu.:      81   3rd Qu.: 9.000   3rd Qu.:     0.00   3rd Qu.:      4.0
##  Max.   :12145792   Max.   :52.000   Max.   :265272.00   Max.   :1510592.0
##                     NA's   :14724
##  forecast_6_month    forecast_9_month    sales_1_month      sales_3_month
##  Min.   :      0.0   Min.   :      0.0   Min.   :     0.0   Min.   :      0.0
##  1st Qu.:      0.0   1st Qu.:      0.0   1st Qu.:     0.0   1st Qu.:      0.0
##  Median :      0.0   Median :      0.0   Median :     0.0   Median :      1.0
##  Mean   :    348.8   Mean   :    508.3   Mean   :    51.5   Mean   :    172.1
##  3rd Qu.:     12.0   3rd Qu.:     20.0   3rd Qu.:     4.0   3rd Qu.:     14.0
##  Max.   :2157024.0   Max.   :3162260.0   Max.   :349620.0   Max.   :1099852.0
##
##  sales_6_month       sales_9_month        min_bank        potential_issue
##  Min.   :      0.0   Min.   :      0   Min.   :     0.0   No :241993
##  1st Qu.:      0.0   1st Qu.:      0   1st Qu.:     0.0   Yes:    82
##  Median :      2.0   Median :      4   Median :     0.0
##  Mean   :    340.4   Mean   :    512   Mean   :    52.8
##  3rd Qu.:     30.0   3rd Qu.:     46   3rd Qu.:     3.0
##  Max.   :2103389.0   Max.   :3195211   Max.   :303713.0
##
##  pieces_past_due    perf_6_month_avg  perf_12_month_avg  local_bo_qty
##  Min.   :    0.00   Min.   :-99.000   Min.   :-99.000   Min.   :   0.000
##  1st Qu.:    0.00   1st Qu.:  0.630   1st Qu.:  0.660   1st Qu.:   0.000
##  Median :    0.00   Median :  0.820   Median :  0.810   Median :   0.000
##  Mean   :    1.82   Mean   : -7.094   Mean   : -6.632   Mean   :   0.844
##  3rd Qu.:    0.00   3rd Qu.:  0.960   3rd Qu.:  0.950   3rd Qu.:   0.000
##  Max.   :79964.00   Max.   :  1.000   Max.   :  1.000   Max.   :6232.000
##
##  deck_risk    oe_constraint ppap_risk    stop_auto_buy rev_stop
##  No :194105   No :242028    No :213357   No :  9458    No :241967
##  Yes: 47970   Yes:    47    Yes: 28718   Yes:232617    Yes:   108
##
##
##
##
##
##  went_on_backorder
##  No :239387
##  Yes:  2688
##
##
##
##
##
##
```

```
p2 = ggplot(testData, aes(went_on_backorder)) + geom_bar() + ggtitle("After Cleanin
g")

grid.arrange(p1,p2, ncol = 2)
```




```
# rows and cols with all NA values and unsed levels are removed from the data set
```

Outlier Detection

```
## TRAINING DATA
continuousVariables <- select_if(trainData,is.numeric)
#continuousVariables <- as.data.frame(colnames(continuousVariables))
#colnames(continuousVariables) <- c("names")
continuousVariables <- colnames(continuousVariables)
continuousVariables
```

```
##  [1] "national_inv"      "lead_time"         "in_transit_qty"
##  [4] "forecast_3_month"  "forecast_6_month"  "forecast_9_month"
##  [7] "sales_1_month"     "sales_3_month"     "sales_6_month"
## [10] "sales_9_month"     "min_bank"          "pieces_past_due"
## [13] "perf_6_month_avg"  "perf_12_month_avg" "local_bo_qty"
```

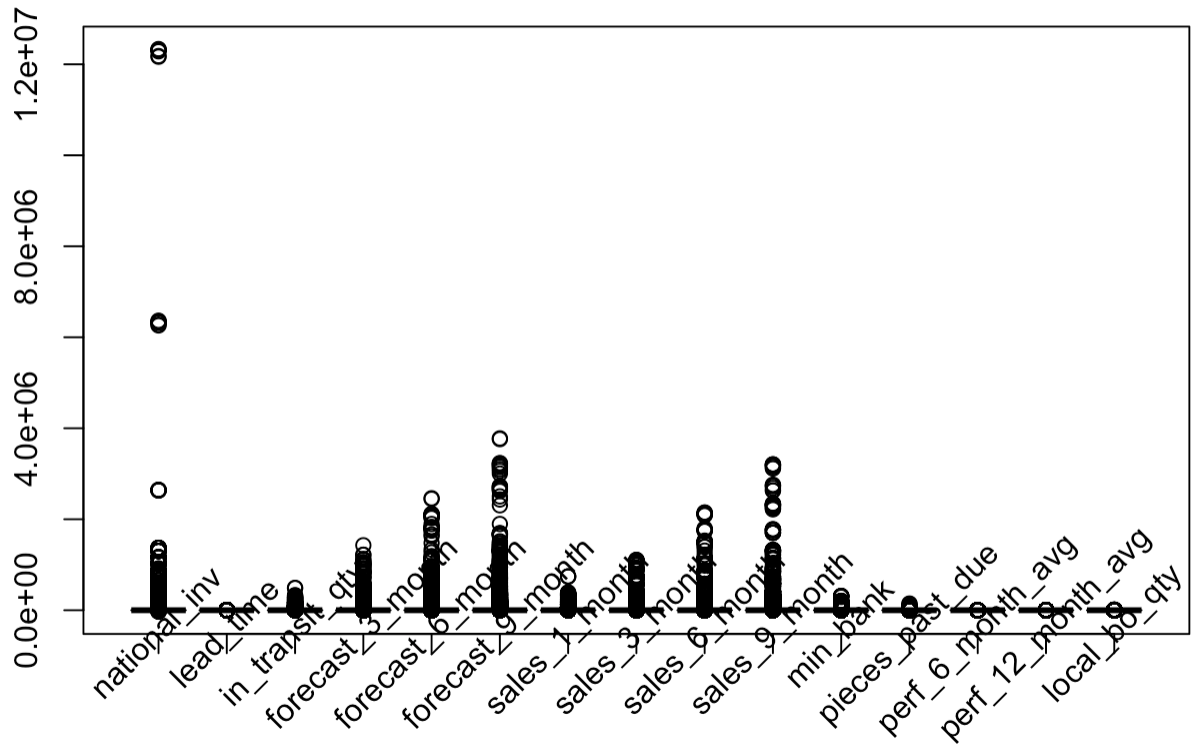
```
library(robustHD)
```

```
## Loading required package: perry
```

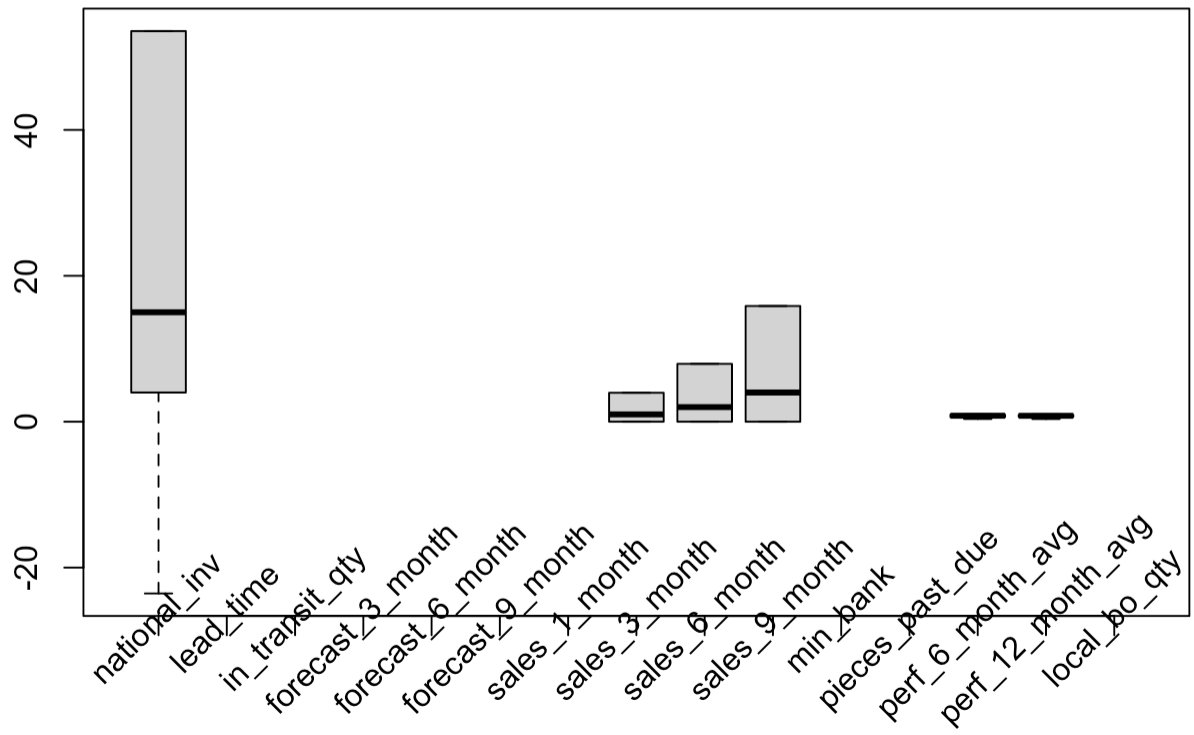
```
## Loading required package: parallel
```

```
## Loading required package: robustbase
```

```
z = trainData
z[,continuousVariables] <- sapply(z[,continuousVariables], function(x) winsorize(x, t
hreshold =1.5, method = "zscore", robust = TRUE))

#par(mfrow=c(1,2))
p1 = boxplot(trainData[,continuousVariables],  xaxt = "n")
tick <- seq_along(p1$names)
axis(1, at = tick, labels = FALSE)
text(tick, par("usr")[3] - 0.45, p1$names, srt = 45, xpd = TRUE)
```



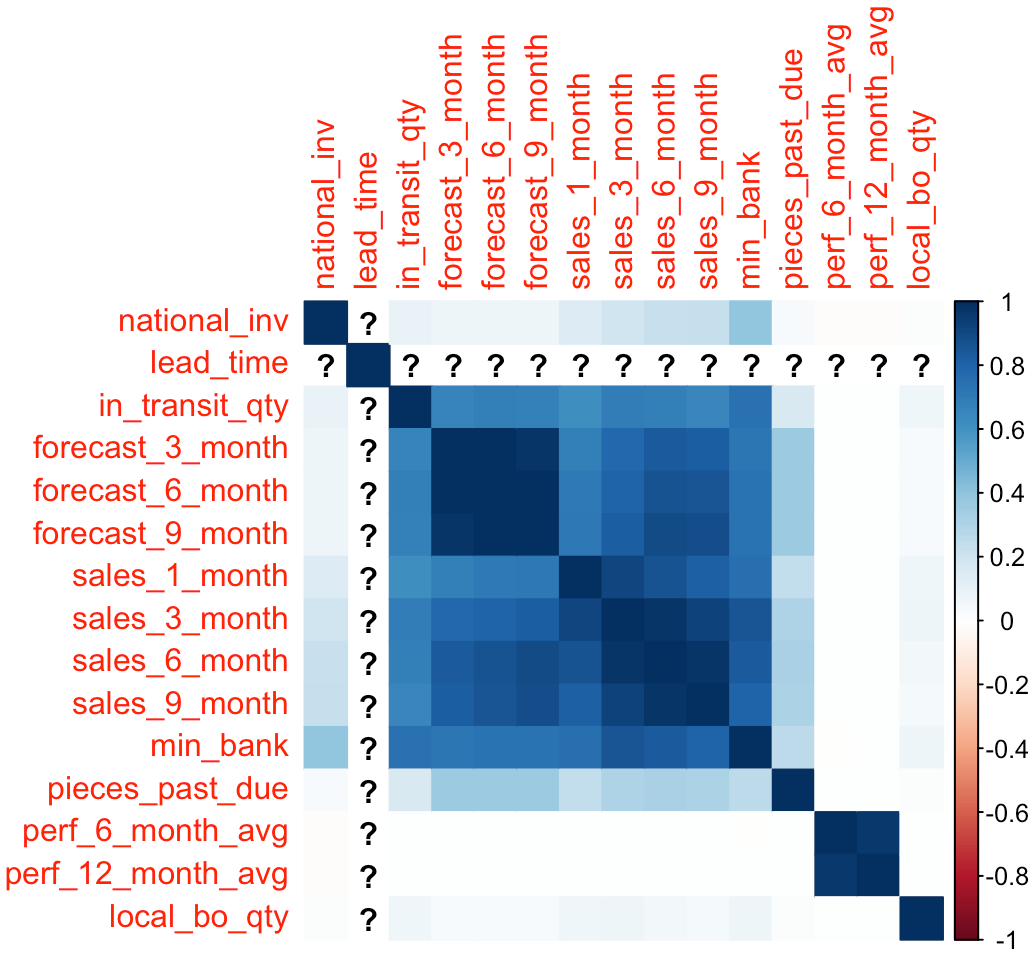
```
p2 = boxplot(z[,continuousVariables], xaxt = "n")
tick <- seq_along(p2$names)
axis(1, at = tick, labels = FALSE)
text(tick, par("usr")[3] - 0.45, p2$names, srt = 45, xpd = TRUE)
```




```
c = cor(trainData[,continuousVariables])

library('corrplot')
```

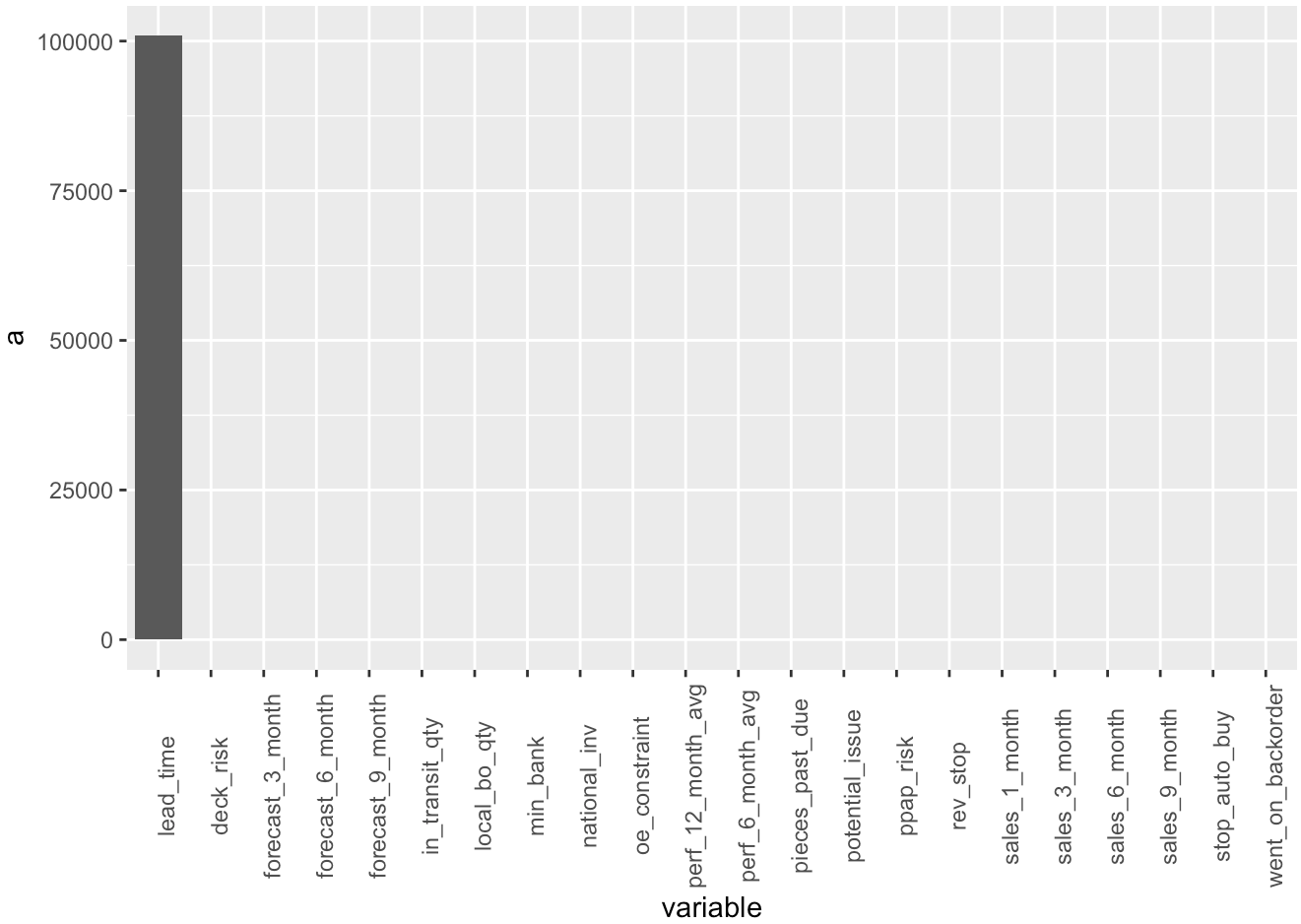
```
## corrplot 0.92 loaded
```

```
corrplot(c, method="color")
```



Data Imputation (solve missing values problem)

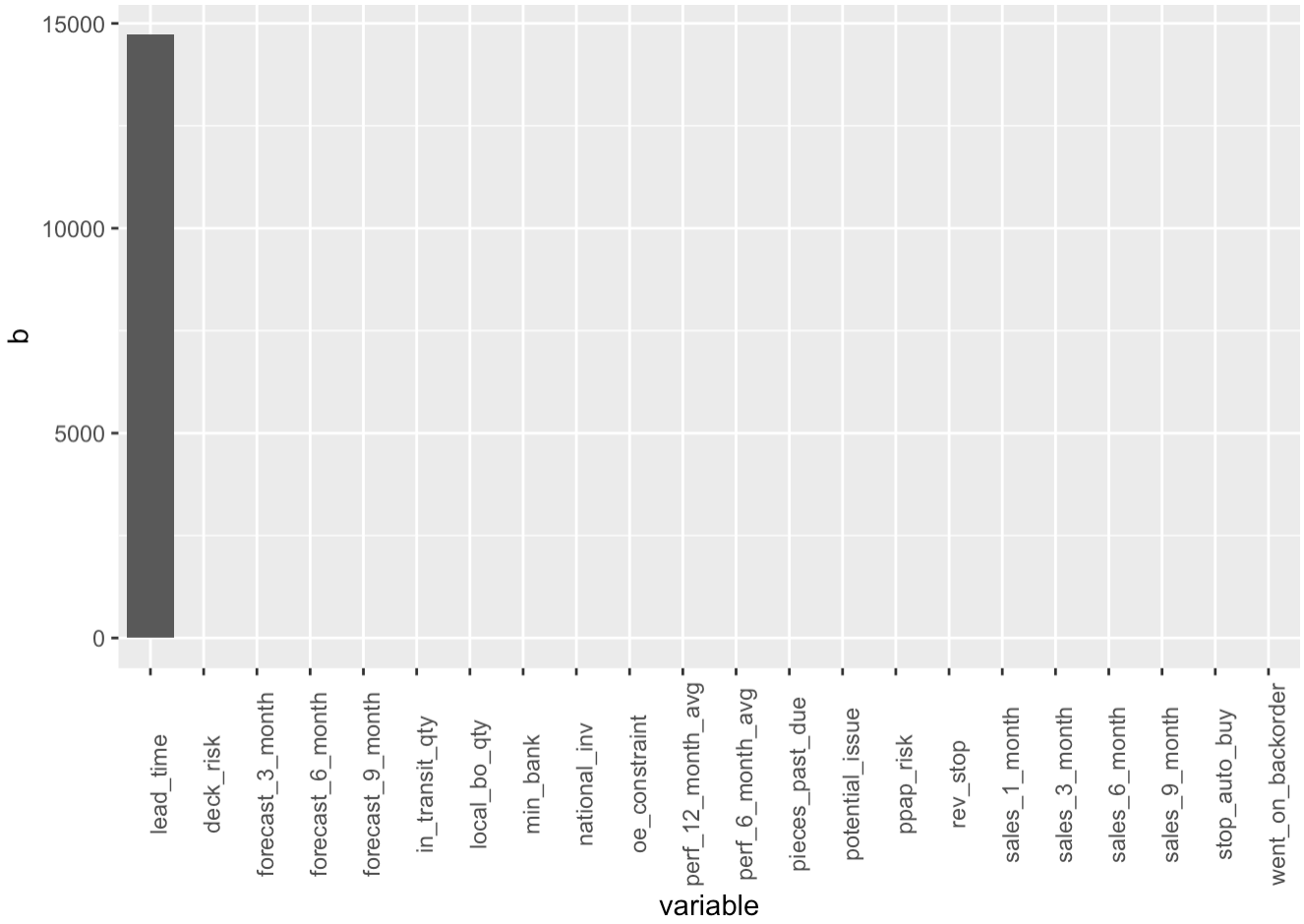
```
# TRAINING DATA

a<-sapply(trainData,function(x)sum(is.na(x)))

data.frame(a,variable=colnames(trainData))%>%
  ggplot(aes(reorder(variable,-a),a))+
    geom_bar(stat="identity")+
    labs(x="variable")+
    theme(axis.text.x = element_text(angle=90))
```




```
# it is seen that the column "lead_time" has a lot of missing values
# therefore we replace it with mean value of the column lead_time

trainData$lead_time[is.na(trainData$lead_time)] = mean(trainData$lead_time, na.rm=TRU
E)

summary(trainData)
```

```
##   national_inv        lead_time      in_transit_qty     forecast_3_month
##  Min.   :  -27256   Min.   : 0.000   Min.   :     0.0   Min.   :      0.0
##  1st Qu.:       4   1st Qu.: 4.000   1st Qu.:     0.0   1st Qu.:      0.0
##  Median :      15   Median : 8.000   Median :     0.0   Median :      0.0
##  Mean   :     496   Mean   : 7.872   Mean   :    44.1   Mean   :    178.1
##  3rd Qu.:      80   3rd Qu.: 8.000   3rd Qu.:     0.0   3rd Qu.:      4.0
##  Max.   :12334404   Max.   :52.000   Max.   :489408.0   Max.   :1427612.0
##  forecast_6_month  forecast_9_month  sales_1_month      sales_3_month
##  Min.   :      0   Min.   :      0   Min.   :     0.0   Min.   :      0
##  1st Qu.:      0   1st Qu.:      0   1st Qu.:     0.0   1st Qu.:      0
##  Median :      0   Median :      0   Median :     0.0   Median :      1
##  Mean   :    345   Mean   :    506   Mean   :    55.9   Mean   :    175
##  3rd Qu.:     12   3rd Qu.:     20   3rd Qu.:     4.0   3rd Qu.:     15
##  Max.   :2461360   Max.   :3777304   Max.   :741774.0   Max.   :1105478
##  sales_6_month       sales_9_month        min_bank         potential_issue
##  Min.   :      0.0   Min.   :      0   Min.   :     0.00   No :1686953
##  1st Qu.:      0.0   1st Qu.:      0   1st Qu.:     0.00   Yes:    907
##  Median :      2.0   Median :      4   Median :     0.00
##  Mean   :    341.7   Mean   :    525   Mean   :    52.77
##  3rd Qu.:     31.0   3rd Qu.:     47   3rd Qu.:     3.00
##  Max.   :2146625.0   Max.   :3205172   Max.   :313319.00
##  pieces_past_due     perf_6_month_avg  perf_12_month_avg  local_bo_qty
##  Min.   :     0.00   Min.   :-99.000   Min.   :-99.000   Min.   :    0.000
##  1st Qu.:     0.00   1st Qu.:  0.630   1st Qu.:  0.660   1st Qu.:    0.000
##  Median :     0.00   Median :  0.820   Median :  0.810   Median :    0.000
##  Mean   :     2.04   Mean   : -6.872   Mean   : -6.438   Mean   :    0.626
##  3rd Qu.:     0.00   3rd Qu.:  0.970   3rd Qu.:  0.950   3rd Qu.:    0.000
##  Max.   :146496.00   Max.   :  1.000   Max.   :  1.000   Max.   :12530.000
##  deck_risk     oe_constraint ppap_risk     stop_auto_buy rev_stop
##  No :1300377   No :1687615   No :1484026   No :  61086   No :1687129
##  Yes: 387483   Yes:    245   Yes: 203834   Yes:1626774   Yes:    731
##
##
##
##
##  went_on_backorder
##  No :1676567
##  Yes:  11293
##
##
##
##
```

```
# TESTING DATA

b<-sapply(testData,function(x)sum(is.na(x)))

data.frame(a,variable=colnames(testData))%>%
  ggplot(aes(reorder(variable,-b),b))+
    geom_bar(stat="identity")+
    labs(x="variable")+
    theme(axis.text.x = element_text(angle=90))
```




```
# it is seen that the column "lead_time" has a lot of missing values
# therefore we replace it with mean value of the column lead_time

testData$lead_time[is.na(testData$lead_time)] = mean(testData$lead_time, na.rm=TRUE)

summary(testData)
```

```
##   national_inv        lead_time      in_transit_qty      forecast_3_month
##  Min.   :  -25414   Min.   : 0.000   Min.   :     0.00   Min.   :      0.0
##  1st Qu.:       4   1st Qu.: 4.000   1st Qu.:     0.00   1st Qu.:      0.0
##  Median :      15   Median : 8.000   Median :     0.00   Median :      0.0
##  Mean   :     500   Mean   : 7.923   Mean   :    36.18   Mean   :    181.5
##  3rd Qu.:      81   3rd Qu.: 8.000   3rd Qu.:     0.00   3rd Qu.:      4.0
##  Max.   :12145792   Max.   :52.000   Max.   :265272.00   Max.   :1510592.0
##  forecast_6_month    forecast_9_month    sales_1_month      sales_3_month
##  Min.   :      0.0   Min.   :      0.0   Min.   :     0.0   Min.   :      0.0
##  1st Qu.:      0.0   1st Qu.:      0.0   1st Qu.:     0.0   1st Qu.:      0.0
##  Median :      0.0   Median :      0.0   Median :     0.0   Median :      1.0
##  Mean   :    348.8   Mean   :    508.3   Mean   :    51.5   Mean   :    172.1
##  3rd Qu.:     12.0   3rd Qu.:     20.0   3rd Qu.:     4.0   3rd Qu.:     14.0
##  Max.   :2157024.0   Max.   :3162260.0   Max.   :349620.0   Max.   :1099852.0
##  sales_6_month       sales_9_month        min_bank        potential_issue
##  Min.   :      0.0   Min.   :      0   Min.   :     0.0   No :241993
##  1st Qu.:      0.0   1st Qu.:      0   1st Qu.:     0.0   Yes:    82
##  Median :      2.0   Median :      4   Median :     0.0
##  Mean   :    340.4   Mean   :    512   Mean   :    52.8
##  3rd Qu.:     30.0   3rd Qu.:     46   3rd Qu.:     3.0
##  Max.   :2103389.0   Max.   :3195211   Max.   :303713.0
##  pieces_past_due    perf_6_month_avg  perf_12_month_avg  local_bo_qty
##  Min.   :    0.00   Min.   :-99.000   Min.   :-99.000   Min.   :   0.000
##  1st Qu.:    0.00   1st Qu.:  0.630   1st Qu.:  0.660   1st Qu.:   0.000
##  Median :    0.00   Median :  0.820   Median :  0.810   Median :   0.000
##  Mean   :    1.82   Mean   : -7.094   Mean   : -6.632   Mean   :   0.844
##  3rd Qu.:    0.00   3rd Qu.:  0.960   3rd Qu.:  0.950   3rd Qu.:   0.000
##  Max.   :79964.00   Max.   :  1.000   Max.   :  1.000   Max.   :6232.000
##  deck_risk    oe_constraint ppap_risk    stop_auto_buy rev_stop
##  No :194105   No :242028    No :213357   No :  9458    No :241967
##  Yes: 47970   Yes:    47    Yes: 28718   Yes:232617    Yes:   108
##
##
##
##
##  went_on_backorder
##  No :239387
##  Yes:  2688
##
##
##
##
```

CONVERTING ALL CHARCHTER VALUES TO NUMERIC

```
#function to map yes and no values to 1 and 0
Mapvalue <- function(x) {
  plyr::mapvalues(x, from = c("Yes", "No"), to = c(1,0))
}

# TRAINING DATA
categoricalData1<-select_if(trainData,is.factor)
categoricalData1[c(1:ncol(categoricalData1))] <- lapply(categoricalData1[c(1:ncol(cat
egoricalData1))], Mapvalue)
NumericData1<-select_if(trainData,is.numeric)
trainData <- cbind(NumericData1, categoricalData1)
head(trainData)
```

```
##   national_inv lead_time in_transit_qty forecast_3_month forecast_6_month
## 1            0  7.872267              0                0                0
## 2            2  9.000000              0                0                0
## 3            2  7.872267              0                0                0
## 4            7  8.000000              0                0                0
## 5            8  7.872267              0                0                0
## 6           13  8.000000              0                0                0
##   forecast_9_month sales_1_month sales_3_month sales_6_month sales_9_month
## 1                0             0             0             0             0
## 2                0             0             0             0             0
## 3                0             0             0             0             0
## 4                0             0             0             0             0
## 5                0             0             0             0             4
## 6                0             0             0             0             0
##   min_bank pieces_past_due perf_6_month_avg perf_12_month_avg local_bo_qty
## 1        0               0           -99.00            -99.00            0
## 2        0               0             0.99              0.99            0
## 3        0               0           -99.00            -99.00            0
## 4        1               0             0.10              0.13            0
## 5        2               0           -99.00            -99.00            0
## 6        0               0             0.82              0.87            0
##   potential_issue deck_risk oe_constraint ppap_risk stop_auto_buy rev_stop
## 1               0         0             0         0             1        0
## 2               0         0             0         0             1        0
## 3               0         1             0         0             1        0
## 4               0         0             0         0             1        0
## 5               0         1             0         0             1        0
## 6               0         0             0         0             1        0
##   went_on_backorder
## 1                 0
## 2                 0
## 3                 0
## 4                 0
## 5                 0
## 6                 0
```

```
# TESTING DATA
categoricalData2<-select_if(testData,is.factor)
categoricalData2[c(1:ncol(categoricalData2))] <- lapply(categoricalData2[c(1:ncol(cat
egoricalData2))], Mapvalue)
NumericData2<-select_if(testData,is.numeric)
testData <- cbind(NumericData2, categoricalData2)
head(testData)
```

```
##   national_inv lead_time in_transit_qty forecast_3_month forecast_6_month
## 1           62  7.923018              0                0                0
## 2            9  7.923018              0                0                0
## 3           17  8.000000              0                0                0
## 4            9  2.000000              0                0                0
## 5            2  8.000000              0                0                0
## 6           15  2.000000              0                0                0
##   forecast_9_month sales_1_month sales_3_month sales_6_month sales_9_month
## 1                0             0             0             0             0
## 2                0             0             0             0             0
## 3                0             0             0             0             0
## 4                0             0             0             0             2
## 5                0             0             0             0             0
## 6                0             0             0             1             2
##   min_bank pieces_past_due perf_6_month_avg perf_12_month_avg local_bo_qty
## 1        1               0           -99.00            -99.00            0
## 2        1               0           -99.00            -99.00            0
## 3        0               0             0.92              0.95            0
## 4        0               0             0.78              0.75            0
## 5        0               0             0.54              0.71            0
## 6        0               0             0.37              0.68            0
##   potential_issue deck_risk oe_constraint ppap_risk stop_auto_buy rev_stop
## 1               0         1             0         0             1        0
## 2               0         0             0         1             0        0
## 3               0         0             0         0             1        0
## 4               0         0             0         1             1        0
## 5               0         0             0         0             1        0
## 6               0         0             0         0             1        0
##   went_on_backorder
## 1                 0
## 2                 0
## 3                 0
## 4                 0
## 5                 0
## 6                 0
```

SAMPLING

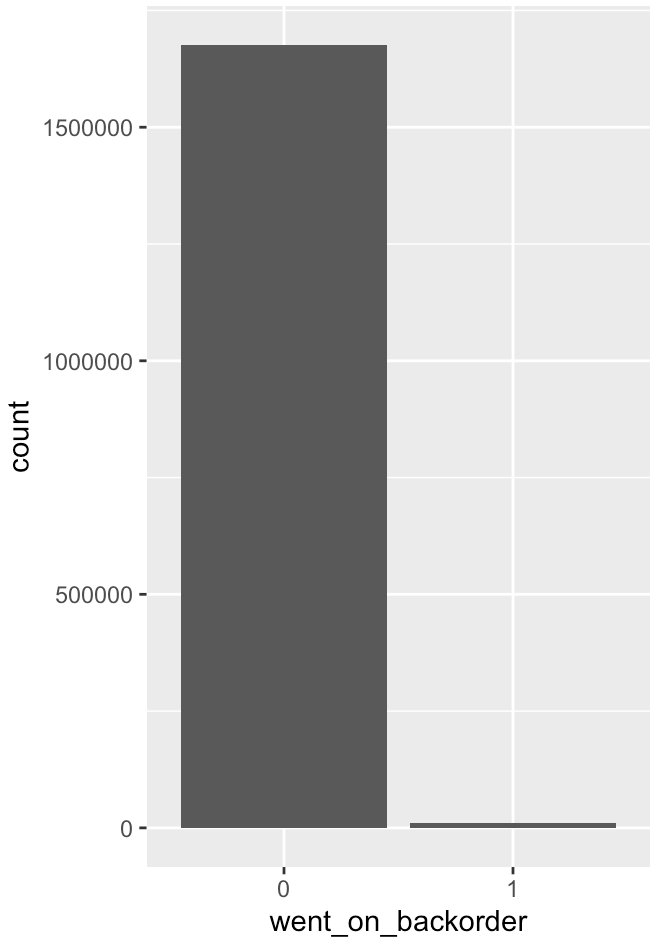
```
# It is clearly seen that the outcome variable 'went_on_backorder' is biased, therefo
re it is necessary to oversample the data so as to get a better accuracy and fit on t
he model

sampledTrainData <- ovun.sample(went_on_backorder~., data = trainData, method = "bot
h")$data
table(trainData$went_on_backorder)
```

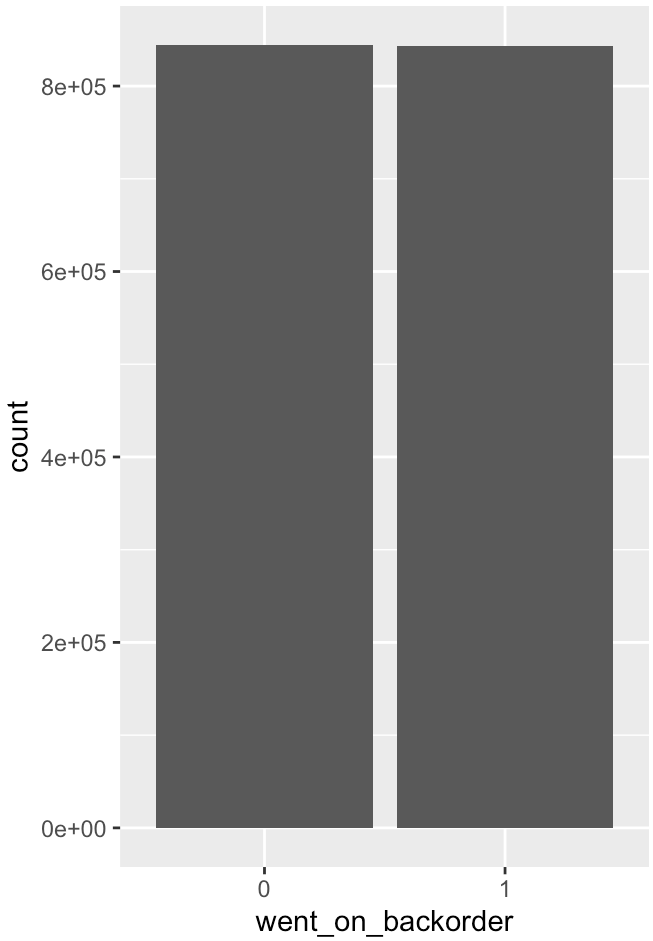
```
##
##       0       1
## 1676567   11293
```

```
table(sampledTrainData$went_on_backorder)
```

```
##
##      0      1
## 844540 843320
```

```
grid.arrange(
  ggplot(trainData, aes(went_on_backorder)) + geom_bar() ,
  ggplot(sampledTrainData, aes(went_on_backorder)) + geom_bar(),
  ncol = 2)
```





MODELING

```
# LOGISTIC REGRESSION
fit.glm1 <- caret::train(went_on_backorder~ lead_time +sales_1_month+ national_inv  +
sales_3_month + forecast_9_month , data=trainData, method="glm",family= binomial)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
varImp(fit.glm1)
```

```
## glm variable importance
##
##                   Overall
## forecast_9_month  100.00
## sales_3_month      64.41
## sales_1_month      33.36
## lead_time          32.43
## national_inv        0.00
```

```
fit.glm1$results
```

```
##   parameter  Accuracy        Kappa   AccuracySD      KappaSD
## 1      none 0.9932977 6.357054e-05 8.968729e-05 0.0002133371
```

```
fit.glm2 <- caret::train(went_on_backorder~ lead_time + national_inv , data=sampledTr
ainData, method="glm",family= binomial)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
varImp(fit.glm2)
```

```
## glm variable importance
##
##              Overall
## national_inv     100
## lead_time          0
```

```
fit.glm2$results
```

```
##   parameter  Accuracy     Kappa   AccuracySD     KappaSD
## 1      none 0.6472234 0.2946088 0.0009411558 0.001730002
```

```
# DECISION TREE
fit.dt1 <- caret::train(went_on_backorder~., data=sampledTrainData, method="rpart")
varImp(fit.dt1)
```

```
## rpart variable importance
##
##   only 20 most important variables shown (out of 21)
##
##                   Overall
## national_inv      100.000
## forecast_3_month   96.365
## forecast_6_month   94.135
## forecast_9_month   90.453
## sales_3_month      29.404
## in_transit_qty     19.452
## min_bank            9.344
## sales_9_month       9.207
## sales_6_month       8.744
## rev_stop1           0.000
## pieces_past_due     0.000
## stop_auto_buy1      0.000
## perf_12_month_avg   0.000
## potential_issue1    0.000
## deck_risk1          0.000
## sales_1_month       0.000
## perf_6_month_avg    0.000
## local_bo_qty        0.000
## ppap_risk1          0.000
## oe_constraint1      0.000
```

```
fit.dt1$results
```

```
##            cp  Accuracy     Kappa  AccuracySD     KappaSD
## 1 0.008278589 0.8267304 0.6534521 0.004134351 0.008273373
## 2 0.093031115 0.7976739 0.5953765 0.023731698 0.047432262
## 3 0.554574776 0.6328850 0.2659519 0.141316659 0.282520032
```

```
fit.dt2 <- caret::train(went_on_backorder~ lead_time +sales_9_month+ national_inv  +
sales_6_month + forecast_3_month , data=sampledTrainData, method="rpart")
varImp(fit.dt2)
```

```
## rpart variable importance
##
##                  Overall
## national_inv      100.00
## forecast_3_month   99.18
## sales_6_month      31.23
## sales_9_month      28.99
## lead_time           0.00
```

```
fit.dt2$results
```

```
##            cp  Accuracy     Kappa  AccuracySD     KappaSD
## 1 0.008461201 0.8290193 0.6580340 0.004269932 0.008542692
## 2 0.093031115 0.8016304 0.6032760 0.023674124 0.047323880
## 3 0.554574776 0.5887789 0.1772502 0.131733778 0.263712602
```

RESULTS

```
results <- resamples(list(GLM=fit.glm1,CART=fit.dt2))
summary(results)
```

```
##
## Call:
## summary.resamples(object = results)
##
## Models: GLM, CART
## Number of resamples: 25
##
## Accuracy
##           Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## GLM  0.9931247 0.9932249 0.9932940 0.9932977 0.9933665 0.9934782    0
## CART 0.8231822 0.8240413 0.8318696 0.8290193 0.8324967 0.8336328    0
##
## Kappa
##                Min.       1st Qu.    Median         Mean   3rd Qu.         Max.
## GLM  -6.723814e-05 -3.221929e-06 0.0000000 6.357054e-05 0.0000000 0.0009365312
## CART  6.463447e-01  6.480867e-01 0.6637345 6.580340e-01 0.6649945 0.6672657609
##      NA's
## GLM     0
## CART    0
```

```
summary(fit.glm1)
```

```
##
## Call:
## NULL
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
##  -8.49    0.00    0.00    0.00    8.49
##
## Coefficients:
##                    Estimate Std. Error    z value Pr(>|z|)
## (Intercept)      -4.013e+15  7.878e+04 -5.094e+10   <2e-16 ***
## lead_time         8.920e+11  7.550e+03  1.181e+08   <2e-16 ***
## sales_1_month     8.456e+09  6.971e+01  1.213e+08   <2e-16 ***
## national_inv     -1.541e+07  1.808e+00 -8.522e+06   <2e-16 ***
## sales_3_month    -7.461e+09  3.297e+01 -2.263e+08   <2e-16 ***
## forecast_9_month  2.315e+09  6.678e+00  3.466e+08   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 135599  on 1687859  degrees of freedom
## Residual deviance: 814154  on 1687854  degrees of freedom
## AIC: 814166
##
## Number of Fisher Scoring iterations: 19
```

```
caret::confusionMatrix(sampledTrainData$went_on_backorder,predict(fit.glm1),positive
= "1")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction      0      1
##          0 844540      0
##          1 843319      1
##
##                Accuracy : 0.5004
##                  95% CI : (0.4996, 0.5011)
##     No Information Rate : 1
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 1.000e+00
##             Specificity : 5.004e-01
##          Pos Pred Value : 1.186e-06
##          Neg Pred Value : 1.000e+00
##              Prevalence : 5.925e-07
##          Detection Rate : 5.925e-07
##    Detection Prevalence : 4.996e-01
##       Balanced Accuracy : 7.502e-01
##
##        'Positive' Class : 1
##
```

Prediction

```
test.glm <- predict(fit.glm1, newdata = testData)
Conflog<-confusionMatrix(data = test.glm , reference =testData$went_on_backorder ,pos
itive = "1")
Conflog
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction      0      1
##          0 239386   2688
##          1      1      0
##
##                Accuracy : 0.9889
##                  95% CI : (0.9885, 0.9893)
##     No Information Rate : 0.9889
##     P-Value [Acc > NIR] : 0.5129
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.000e+00
##             Specificity : 1.000e+00
##          Pos Pred Value : 0.000e+00
##          Neg Pred Value : 9.889e-01
##              Prevalence : 1.110e-02
##          Detection Rate : 0.000e+00
##    Detection Prevalence : 4.131e-06
##       Balanced Accuracy : 5.000e-01
##
##        'Positive' Class : 1
##
```

```
caret::confusionMatrix(sampledTrainData$went_on_backorder,predict(fit.dt1),positive =
"1")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction      0      1
##          0 713914 130626
##          1 166555 676765
##
##                Accuracy : 0.8239
##                  95% CI : (0.8234, 0.8245)
##     No Information Rate : 0.5216
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6478
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.8382
##             Specificity : 0.8108
##          Pos Pred Value : 0.8025
##          Neg Pred Value : 0.8453
##              Prevalence : 0.4784
##          Detection Rate : 0.4010
##    Detection Prevalence : 0.4996
##       Balanced Accuracy : 0.8245
##
##        'Positive' Class : 1
##
```

```
test.DT <- predict(fit.dt1, newdata = testData)
#-- Confusion matrix of test data for Decision tree
Confdt<-confusionMatrix(data = test.DT , reference =testData$went_on_backorder ,posit
ive = "1")
Confdt
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction      0      1
##          0 204320    625
##          1  35067   2063
##
##                Accuracy : 0.8526
##                  95% CI : (0.8511, 0.854)
##     No Information Rate : 0.9889
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.0847
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.767485
##             Specificity : 0.853513
##          Pos Pred Value : 0.055562
##          Neg Pred Value : 0.996950
##              Prevalence : 0.011104
##          Detection Rate : 0.008522
##    Detection Prevalence : 0.153382
##       Balanced Accuracy : 0.810499
##
##        'Positive' Class : 1
##
```

Result of Predicted Model

```
Results1<-cbind(Conflog$byClass,Confdt$byClass)
Results1
```

```
##                             [,1]        [,2]
## Sensitivity          0.000000e+00 0.767485119
## Specificity          9.999958e-01 0.853513349
## Pos Pred Value       0.000000e+00 0.055561541
## Neg Pred Value       9.888960e-01 0.996950401
## Precision            0.000000e+00 0.055561541
## Recall               0.000000e+00 0.767485119
## F1                            NaN 0.103621478
## Prevalence           1.110400e-02 0.011103997
## Detection Rate       0.000000e+00 0.008522152
## Detection Prevalence 4.130951e-06 0.153382216
## Balanced Accuracy    4.999979e-01 0.810499234
```