R Notebook

Code ▪

Hide

```
df <- read.csv( "C:/Users/sach/Downloads/NSE-Tata-Global-Beverages-Limited - NSE-Tata-Global-Beve
rages-Limited.csv" )
View(df)
```

Hide

```
summary(df)
```

```
     Date                Open            High            Low             Last           Close
Total.Trade.Quantity
 Length:1235        Min.   :103.0   Min.   :104.6   Min.   :100.0   Min.   :102.6   Min.   :102.
7   Min.   :  100180
 Class :character   1st Qu.:137.6   1st Qu.:138.9   1st Qu.:135.2   1st Qu.:137.2   1st Qu.:137.
2   1st Qu.: 1284482
 Mode  :character   Median :151.5   Median :153.2   Median :149.5   Median :151.2   Median :151.
1   Median : 1964885
                    Mean   :169.0   Mean   :171.4   Mean   :166.4   Mean   :168.7   Mean   :168.
7   Mean   : 2604151
                    3rd Qu.:169.0   3rd Qu.:172.3   3rd Qu.:166.7   3rd Qu.:169.1   3rd Qu.:169.
5   3rd Qu.: 3095788
                    Max.   :327.7   Max.   :328.8   Max.   :321.6   Max.   :325.9   Max.   :325.
8   Max.   :29191015
 Turnover..Lacs.
 Min.   :  128
 1st Qu.: 1801
 Median : 3069
 Mean   : 4843
 3rd Qu.: 5853
 Max.   :55755
```

Hide

```
str(df)
```

```
'data.frame':   1235 obs. of  8 variables:
 $ Date                : chr  "2018-10-08" "2018-10-05" "2018-10-04" "2018-10-03" ...
 $ Open                : num  208 217 224 230 235 ...
 $ High                : num  222 219 228 238 235 ...
 $ Low                 : num  207 206 216 226 221 ...
 $ Last                : num  216 210 217 226 230 ...
 $ Close               : num  215 209 218 228 231 ...
 $ Total.Trade.Quantity: int  4642146 3519515 1728786 1708590 1534749 3069914 5082859 2240909 23
49368 3423509 ...
 $ Turnover..Lacs.     : num  10063 7407 3816 3960 3486 ...
```

Hide

```
any(is.na(df))
```

```
[1] FALSE
```

Hide

```
attach(df)
```

```
The following objects are masked from df (pos = 8):

    Close, Date, High, Last, Low, Open, Total.Trade.Quantity, Turnover..Lacs.

The following objects are masked from df (pos = 9):

    Close, Date, High, Last, Low, Open, Total.Trade.Quantity, Turnover..Lacs.

The following objects are masked from df (pos = 16):

    Close, Date, High, Last, Low, Open, Total.Trade.Quantity, Turnover..Lacs.
```

Hide

```
sp<-df[,c( "Open","High","Low","Last","Close","Total.Trade.Quantity" ,"Turnover..Lacs." )]
View(data)
```

Hide

```
sp = data.matrix(df, rownames.force =  NA)
```

Hide

```
sp[,'Total.Trade.Quantity' ] = sp[,'Total.Trade.Quantity' ]/10000
```

Hide

```
sp = as.data.frame(data)
View(sp)
```

Hide

```
library('ggplot2')
library('ggthemes')
library('dplyr')
library('caTools')
library('corrgram')
library('corrplot')
```

Hide

```
set.seed(10)
sample = sample.split(df$Turnover..Lacs., SplitRatio = 0.7)
train = subset(sp, sample = TRUE)
test = subset(sp, sample = TRUE)
```

Hide

```
cor(sp)
```

```
                        Date      Open      High       Low      Last     Close Total.Trade.Qua
ntity Turnover..Lacs.
Date                1.0000000 0.6120677 0.6145783 0.6112309 0.6129158 0.6128861            0.20
97206       0.3641588
Open                0.6120677 1.0000000 0.9989565 0.9987757 0.9976617 0.9977038            0.36
75027       0.5870260
High                0.6145783 0.9989565 1.0000000 0.9987276 0.9991297 0.9991592            0.38
87983       0.6059070
Low                 0.6112309 0.9987757 0.9987276 1.0000000 0.9990080 0.9990648            0.36
16948       0.5824463
Last                0.6129158 0.9976617 0.9991297 0.9990080 1.0000000 0.9999628            0.38
12687       0.5995747
Close               0.6128861 0.9977038 0.9991592 0.9990648 0.9999628 1.0000000            0.38
08006       0.5991548
Total.Trade.Quantity 0.2097206 0.3675027 0.3887983 0.3616948 0.3812687 0.3808006            1.00
00000       0.9419757
Turnover..Lacs.     0.3641588 0.5870260 0.6059070 0.5824463 0.5995747 0.5991548            0.94
19757       1.0000000
```

Hide

```
train
```

	D...	Open	High	Low	Last	Close	Total.Trade.Quantity	Turnover..Lacs.
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1235	208.00	222.25	206.85	216.00	215.15	464.2146	10062.83
2	1234	217.00	218.60	205.90	210.25	209.20	351.9515	7407.06
3	1233	223.50	227.80	216.15	217.25	218.20	172.8786	3815.79
4	1232	230.00	237.50	225.75	226.45	227.60	170.8590	3960.27
5	1231	234.55	234.60	221.05	230.30	230.90	153.4749	3486.05
6	1230	234.05	235.95	230.20	233.50	233.75	306.9914	7162.35
7	1229	234.55	236.80	231.10	233.80	233.25	508.2859	11859.95
8	1228	240.00	240.00	232.50	235.00	234.25	224.0909	5248.60
9	1227	233.30	236.75	232.00	236.25	236.10	234.9368	5503.90
10	1226	233.55	239.20	230.75	234.00	233.30	342.3509	7999.55

1-10 of 1,235 rows Previous **1** 2 3 4 5 6 … 100 Next

Hide

```
model1<-lm(Turnover..Lacs.~.,train)
model1
```

```

Call:
lm(formula = Turnover..Lacs. ~ ., data = train)

Coefficients:
         (Intercept)                 Date                 Open                 High
Low
          -5095.2259               0.3467             -45.6668              86.3270
-19.6840
                Last                Close  Total.Trade.Quantity
             48.7313             -43.2741               19.1816
```

Hide

```
summary(model1)
```

```

Call:
lm(formula = Turnover..Lacs. ~ ., data = train)

Residuals:
    Min      1Q  Median      3Q     Max
-6147.9  -527.8    48.3   457.7 14045.0

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          -5095.2259   112.1306 -45.440  <   2e-16 ***
Date                     0.3467     0.1150   3.015  0.00262 **
Open                   -45.6668    21.8191  -2.093  0.03656 *
High                    86.3270    29.1843   2.958  0.00316 **
Low                    -19.6840    23.8982  -0.824  0.41029
Last                    48.7313    72.3182   0.674  0.50054
Close                  -43.2741    77.1332  -0.561  0.57488
Total.Trade.Quantity    19.1816     0.1978  96.991  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1127 on 1227 degrees of freedom
Multiple R-squared:  0.9559,    Adjusted R-squared:  0.9556
F-statistic:  3797 on 7 and 1227 DF,  p-value: < 2.2e-16
```

Hide

```
res<-residuals(model1)
res<-as.data.frame(res)
res
```

	res <dbl>
1	-1005.9318170
2	-777.8529267
3	-1181.1419346
4	-1390.6994702
5	-1210.3119067
6	-470.2287071
7	297.4869929
8	-877.2169308
9	-845.7152124
10	-646.3017176

1-10 of 1,235 rows Previous **1** 2 3 4 5 6 ... 100 Next

Hide

```
Turnover..Lacs..prediction<-predict(model1,test)
result<-cbind(Turnover..Lacs..prediction,test$Turnover..Lacs.)
colnames(result)<-c("pred","real")
result<-as.data.frame(result)
result
```

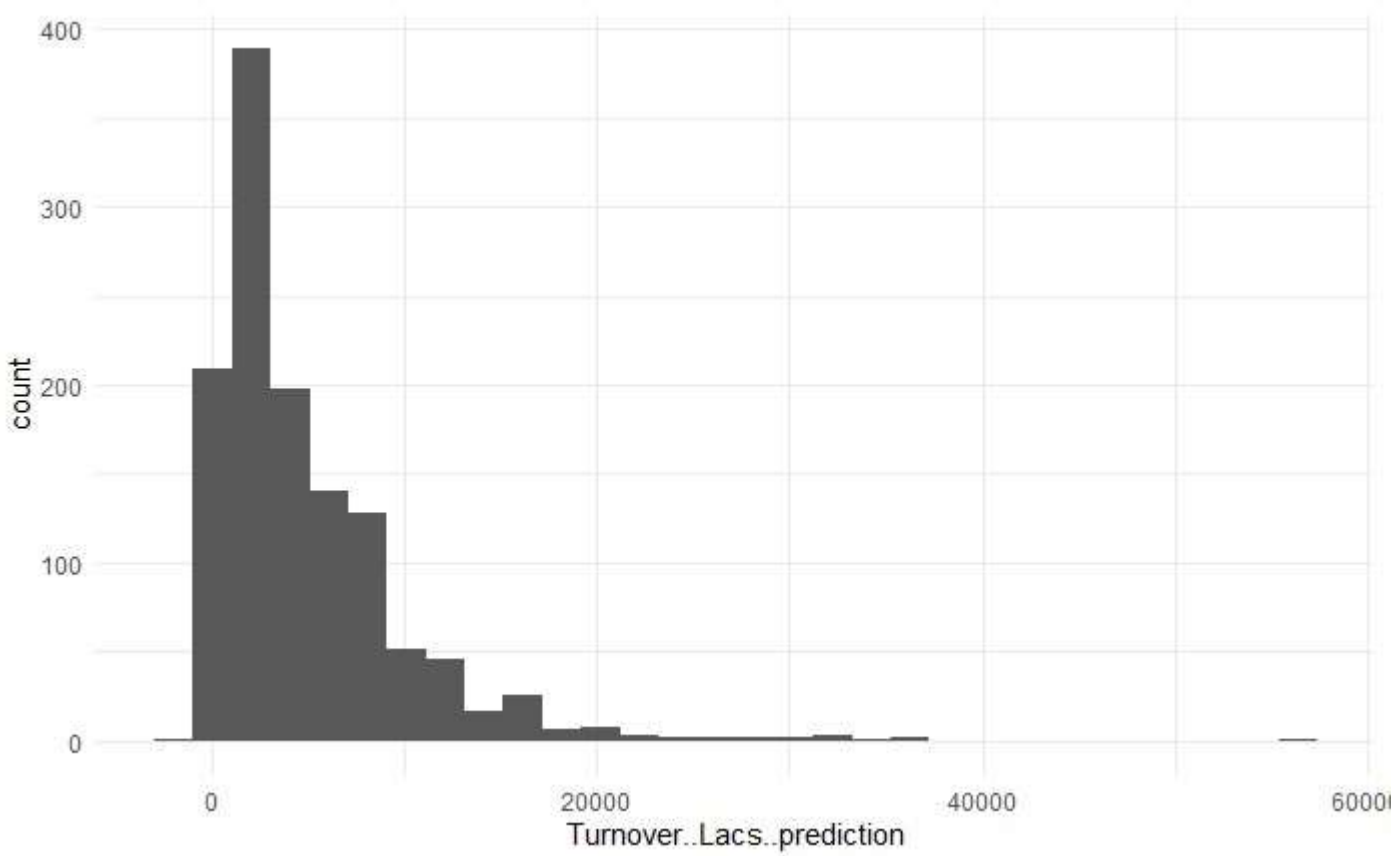
	pred <dbl>	real <dbl>
1	11068.761817	10062.83
2	8184.912927	7407.06
3	4996.931935	3815.79
4	5350.969470	3960.27
5	4696.361907	3486.05
6	7632.578707	7162.35
7	11562.463007	11859.95
8	6125.816931	5248.60
9	6349.615212	5503.90

	pred <dbl>	real <dbl>
10	8645.851718	7999.55

1-10 of 1,235 rows Previous **1** 2 3 4 5 6 ... 100 Next

Hide

```
Turnover..Lacs..prediction = as.data.frame(Turnover..Lacs..prediction)
pl_residuals_test = ggplot(Turnover..Lacs..prediction,aes(Turnover..Lacs..prediction))+geom_hist
ogram()+ theme_minimal()
pl_residuals_test
```



Hide

```
mse<-mean((result$real-result$pred)^ 2)
print(mse)
```

```
[1] 1261634
```

Hide

```
sse<-sum((result$pred-result$real)^ 2)
sst<-sum((result$pred-mean(result$pred))^ 2)
R2<-1-(sse/sst)
R2
```

```
[1] 0.9538305
```

Hide

```
library(plyr)
library(readr)
library(dplyr)
library(caret)
library(ggplot2)
library(repr)
```

Hide

```
dummies <- dummyVars(Turnover..Lacs. ~ ., data = sp)
train_dummies = predict(dummies, newdata = train)
test_dummies = predict(dummies, newdata = test)
print(dim(train_dummies)); print(dim(test_dummies))
```

```
[1] 1235    7
[1] 1235    7
```

Hide

```
library(glmnet)
```

```
package 愀牠glmnet愀牷 was built under R version 4.0.4Loading required package: Matrix
Loaded glmnet 4.1-1
```

Hide

```
x = as.matrix(train_dummies)
y_train = train$Turnover..Lacs.
x_test = as.matrix(test_dummies)
y_test = test$Turnover..Lacs.
lambdas <- 10^seq(2, -3, by = -.1)
ridge_reg = glmnet(x, y_train, nlambda = 25, alpha = 0, family = 'gaussian', lambda
= lambdas)
summary(ridge_reg)
```

```
          Length Class     Mode
a0         51    -none-    numeric
beta      357    dgCMatrix S4
df         51    -none-    numeric
dim         2    -none-    numeric
lambda     51    -none-    numeric
dev.ratio  51    -none-    numeric
nulldev     1    -none-    numeric
npasses     1    -none-    numeric
jerr        1    -none-    numeric
offset      1    -none-    logical
call        7    -none-    call
nobs        1    -none-    numeric
```

Hide

```
cv_ridge <- cv.glmnet(x, y_train, alpha =  0, lambda = lambdas)
optimal_lambda <- cv_ridge$lambda.min
optimal_lambda
```

```
[1] 25.11886
```

Hide

```
# Compute R^2 from true and predicted values
eval_results <-  function(true, predicted, df)  {
SSE <- sum((predicted - true)^ 2)
SST <- sum((true - mean(true))^ 2)
R_square <-  1 - SSE / SST
RMSE = sqrt(SSE/nrow(df))
# Model performance metrics
data.frame(
RMSE = RMSE,
Rsquare = R_square
)
}
# Prediction and evaluation on train data
predictions_train <- predict(ridge_reg, s = optimal_lambda, newx = x)
eval_results(y_train, predictions_train, train)
```

RMSE <dbl>	Rsquare <dbl>
1128.973	0.9554152

1 row

Hide

```
# Prediction and evaluation on test data
predictions_test <- predict(ridge_reg, s = optimal_lambda, newx = x_test)
eval_results(y_test, predictions_test, test)
```

RMSE <dbl>	Rsquare <dbl>
1128.973	0.9554152

1 row