MLR-Code-With-State.R

obite

2021-07-14

df <- read.csv("Dummy Data With State.csv")  
  
# Splitting the dataset into the Training set and Test set  
library (caTools)

## Warning: package 'caTools' was built under R version 4.0.5

set.seed(101)   
sample = sample.split(df$Total.Amount, SplitRatio = .60)  
train = subset(df, sample == TRUE)  
test = subset(df, sample == FALSE)  
  
# use lm() to run a linear regression of Total Amount on all the predictors   
# in the training set.  
model <- lm(Total.Amount ~ ., data = train)  
  
# use options() to ensure numbers are not displayed in scientific notation.  
options(scipen = 999)  
  
summary(model)

##   
## Call:  
## lm(formula = Total.Amount ~ ., data = train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -535307 -160783 -53878 23682 3211184   
##   
## Coefficients: (3 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1249118.90754 351578.51560 -3.553 0.00042 \*\*\*  
## ï..Billing.Country..Dummy. -70239.27646 128379.31759 -0.547 0.58456   
## Billing.State..Dummy. 88983.93166 44328.44255 2.007 0.04529 \*   
## Browser..Dummy. 21123.31588 35888.96985 0.589 0.55643   
## Created.By..Dummy. -23603.04808 28746.15461 -0.821 0.41202   
## Device.Brand..Dummy. 44586.42689 24637.32534 1.810 0.07099 .   
## Device.Type..Dummy. 111452.78373 54257.76041 2.054 0.04052 \*   
## Email.Status..Dummy. 558195.17643 151243.85293 3.691 0.00025 \*\*\*  
## Event.Type..Dummy. NA NA NA NA   
## Order.Source..Dummy. NA NA NA NA   
## Order.Status..Dummy. -42657.39232 28425.29446 -1.501 0.13412   
## Os..Dummy. -97862.25625 50750.51610 -1.928 0.05443 .   
## Os.Version..Dummy. 17101.72659 25498.55991 0.671 0.50275   
## Payment.Method..Dummy. -31990.52576 48483.83816 -0.660 0.50970   
## Segment..Dummy. 102606.02984 35956.56502 2.854 0.00452 \*\*   
## Updated.By..Dummy. NA NA NA NA   
## Discounted.Price -0.12519 0.06888 -1.818 0.06978 .   
## Item.Count 98.30073 209.10484 0.470 0.63850   
## Quantity 76.55723 246.98318 0.310 0.75672   
## Shipping -3.45635 6.93401 -0.498 0.61840   
## Unit.Price 0.39023 0.26683 1.463 0.14428   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 363200 on 462 degrees of freedom  
## Multiple R-squared: 0.1251, Adjusted R-squared: 0.09291   
## F-statistic: 3.886 on 17 and 462 DF, p-value: 0.0000003697

#CODE FOR PREDICTION AND MEASURING ACCURACY  
  
library(forecast)

## Warning: package 'forecast' was built under R version 4.0.5

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

# use predict() to make predictions on a new set.  
pred <- predict(model, test)

## Warning in predict.lm(model, test): prediction from a rank-deficient fit may be  
## misleading

options(scipen=999, digits = 0)  
  
#some.residuals <- test$Total.Amount - pred  
#data.frame("Predicted" = pred, "Actual" = test$Total.Amount,  
# "Residual" = some.residuals)  
  
#options(scipen=999, digits = 3)  
  
# use accuracy() to compute common accuracy measures.  
accuracy(pred, test$Total.Amount)

## ME RMSE MAE MPE MAPE  
## Test set -5549 370820 200650 -822 929