Problem Statement: A German book retailer sent out offers to customers. You job is to build predictor variables and a machine learning model to predict how much each customer who received the contact point will spend. The outcome variable is logtarg. The test set is indicated by is.na(logtarg). The objective is to predict logtarg with minimum RMSE on the test set.

Files Given:

target.csv : The target.csv file has one record for every customer in the training and test sets

orders.csv: The orders.csv file has one record for every item ordered by a customer in the test or training set prior to the contact point.

A screenshot of a cell phone

Description automatically generated

A screenshot of text

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A picture containing clock, meter

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Summary of summarized order

A close up of text on a white background

Description automatically generated

correlation between different parameter

A screenshot of a cell phone

Description automatically generated

A picture containing sitting, photo, old, white

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Overall signifance test

H0: every coefficient is 0

Ha : atleast one coefficient is non zero

as p< 0.05. Rejecting null hypothesis

also by looking at model tof and m are insignificant

So, this model is significant

A screenshot of a cell phone

Description automatically generated

Residual plot

A close up of a map

Description automatically generated

A screenshot of a cell phone

Description automatically generated

==)This model seems multiplicative model as there is a heteroscedasticity and outliers so we should use log log model

<https://www.youtube.com/watch?v=cadV3wXjLRU&feature=youtu.be>