**Linear regression on HPdataset**

Steps:

* First divided whole dataset into training and test set such that training set contains data upto 16th year 4th quarter and testing set contains 17th and 18th year of data. Then run model on it and calculated mse.

It comes to be 9062436171.

* After this taking whole data up to 19th year 1st quarter(19Q1) data as training data as actuals are present for those and taking 19th year 3rd and 4th quarter (19Q2 and 19Q3) data in testing set as actuals value is NULL. And after training model, predicted actuals for 19Q2 and 19Q3 data.
* And then done it for each UID.

**# without scaling**

**Code:**

u<-unique(data2[1])

n<-nrow(u)

df=data.frame(UID=character(0),Pred\_19Q2=numeric(0),Pred\_19Q3=numeric(0))

i=1

while (i<=n){

ntrain<-subset(data2,UID==u[i,] & (Period=='14-Q1'| Period=='14-Q2'| Period=='14-Q3'| Period=='14-Q4'| Period=='15-Q1'| Period=='15-Q2'| Period=='15-Q3'| Period=='15-Q4'| Period=='16-Q1'| Period=='16-Q2'| Period=='16-Q3'| Period=='16-Q4'| Period=='17-Q1'| Period=='17-Q2'| Period=='17-Q3'| Period=='17-Q4'| Period=='18-Q1'| Period=='18-Q2'| Period=='18-Q3'| Period=='18-Q4'| Period=='19-Q1'))

ntest<-subset(data2,UID==u[i,] & (Period=='19-Q2'| Period=='19-Q3'))

model<- lm(Actuals~Market\_Share\_Units+Market\_Share\_Revenue+Market\_Size\_Units+Market\_Size\_Revenue+MEI\_CPI\_Inflation+MEI\_Exports+MEI\_Fixed\_Investment+MEI\_GDP+MEI\_Government\_Growth+MEI\_Imports+MEI\_Industrial\_Production+MEI\_Merchandise\_Exports+MEI\_Merchandise\_Imports+MEI\_Nominal\_Retail\_Sales+MEI\_Private\_Consumption+MEI\_Real\_Retail\_Sales+MEI\_WPI\_Inflation+MEI\_Price\_Index+MEI\_Trade\_GDP\_Ratio+MEI\_Merchandise\_Trade\_GDP\_Ratio+MEI\_Real\_Nominal\_Sales,ntrain,na.action = na.exclude)

sm<-summary(model)

new\_test<-data.frame(ntest[c(12:32)])

pred<-predict(model,newdata=new\_test)

df1<-data.frame(UID=u[i,],Pred\_19Q2=pred[[1]],Pred\_19Q3=pred[[2]])

df<-rbind(df,df1)

i=i+1

}

df

**Output:**

UID Pred\_19Q2 Pred\_19Q3

1 1-V3566F-BPS STM 459675.0185 472505.2423

2 1-10LZ6-8864-OPS HW A4 172621.2412 78931.1898

3 1-10M46-18886-PC Value 12024.3710 -1055.1055

4 Sweden-Print Value 416908.1369 448389.1804

5 1-10LZ6-6794-PC Value 11489.0462 7679.3546

6 1-10LQI-8358-BPS Attach Indirect 6251.1609 3015.1350

7 1-10M99-13159-PC Value 292707.3373 297381.2191

8 1-10LQI-1228-BPS Indirect 222077.8138 294121.8272

9 1-10LQI-14992-BPS Attach Indirect 105663.0150 69673.4404

10 1-10LZ6-9068-BPS Indirect -100753.7242 -10904.2581

11 1-10LQI-15642-PC Value 65986.4227 -131836.3612

12 1-10LQI-22873-OPS HW A4 48685.7498 83863.5074

13 1-10LQI-1385-BPS Indirect -186354.7032 -352188.6173

14 1-10LZ6-6027-OPS HW A4 211938.6379 -136726.7898

15 1-10M46-1749-BPS Indirect 797640.4393 859193.2991

16 1-10M99-7771-BPS Indirect 432842.5184 393707.5472

17 1-10M46-8119-PC Value -8070.8210 -84664.8512

18 1-10LQI-15422-BPS Indirect 90357.8458 170997.4690

19 1-10LZ6-8097-Print Value 805673.3661 1413347.7656

20 1-10M99-20490-BPS Indirect 128914.0991 394169.3120

21 1-10MDH-23535-Print Value 296750.3878 633831.3287

22 1-10M46-8643-PC Value -17792.3001 -264251.3518

23 1-10LQI-21206-OPS HW A4 -177455.1364 2313894.1781

24 1-10M99-12334-PC Value 23553.0331 29308.4833

25 1-10MDH-12319-PC Value -106408.5947 -250914.8148

26 1-10MDH-1735-PC Value -63577.4972 -301513.8753

27 1-10M46-7822-PC Value 15490.2310 40267.9347

28 1-10M46-266-BPS Indirect 2746.5512 14970.3485

29 1-10LZ6-18292-BPS Indirect 65205.3286 29658.8742

30 1-10LQI-11865-BPS STM -417869.9842 246225.4759

31 1-10LZ6-10399-BPS STM 232085.9146 -2254680.3267

32 1-10MDH-21749-BPS Attach Indirect 213965.4085 256364.7053

33 1-10M99-7164-BPS Indirect 550590.3334 915380.6924

34 1-10LQI-17174-OPS HW A3 213229.0266 420229.0920

35 1-10M6L-7017-BPS Indirect 791.4672 10712.1364

36 1-10LQI-3842-PC Value 172975.5782 -20615.8555

37 1-10M46-19931-BPS Attach Indirect 75820.0263 95712.3731

38 1-10M99-17081-OPS HW A4 -94482.1116 -478469.1462

39 1-10M99-5607-PC Value 50831.6150 157021.1047

40 1-10LQI-15441-BPS Indirect -15058.8201 -26747.7412

41 1-10M99-23758-OPS HW A4 10786.3554 29802.4875

42 1-10LT8-22536-BPS Indirect 198957.2635 143001.7349

43 1-10MDH-9435-OPS HW A3 -77665.9477 39403.4827

44 1-10LT8-17950-BPS Indirect -65744.4366 -20938.8415

45 1-10LT8-17471-OPS HW A4 29791.1459 30086.0685

46 1-10LQI-9214-HPS HW -22070.5722 2324.1219

47 1-10M46-10795-BPS Attach Indirect 15199.8519 -34977.4744

48 1-10LQI-1702-BPS STM 226154.8092 384081.0873

49 1-10LZ6-1583-BPS Indirect -34374.2562 -239400.6329

50 1-10M6L-16719-BPS Indirect 169565.9878 316826.8944

51 1-10LT8-14058-PC Value -191833.0655 -332820.3553

52 1-10LZ6-13841-PC Value 38934.1533 57255.9861

53 1-10M46-20972-BPS Indirect 591614.4925 444916.6853

54 1-10LT8-513-PC Value 597397.8702 1585295.3982[ reached 'max' / getOption("max.print") -- omitted 2693 rows ]

**# with scaling**

**Code:**

# for each uid

u<-unique(data2[1])

n<-nrow(u)

df=data.frame(UID=character(0),Pred\_19Q2=numeric(0),Pred\_19Q3=numeric(0))

i=1

while (i<=n){

ntrain<-subset(data2,UID==u[i,] & (Period=='14-Q1'| Period=='14-Q2'| Period=='14-Q3'| Period=='14-Q4'| Period=='15-Q1'| Period=='15-Q2'| Period=='15-Q3'| Period=='15-Q4'| Period=='16-Q1'| Period=='16-Q2'| Period=='16-Q3'| Period=='16-Q4'| Period=='17-Q1'| Period=='17-Q2'| Period=='17-Q3'| Period=='17-Q4'| Period=='18-Q1'| Period=='18-Q2'| Period=='18-Q3'| Period=='18-Q4'| Period=='19-Q1'))

ntest<-subset(data2,UID==u[i,] & (Period=='19-Q2'| Period=='19-Q3'))

ntrain[c(11:32)] <- lapply(ntrain[c(11:32)], function(x) c(scale(x)))

ntest[c(11:32)] <- lapply(ntest[c(11:32)], function(x) c(scale(x)))

ntrain[is.na(ntrain)] <- 0

ntest[is.na(ntest)] <- 0

model<- lm(Actuals~Market\_Share\_Units+Market\_Share\_Revenue+Market\_Size\_Units+Market\_Size\_Revenue+MEI\_CPI\_Inflation+MEI\_Exports+MEI\_Fixed\_Investment+MEI\_GDP+MEI\_Government\_Growth+MEI\_Imports+MEI\_Industrial\_Production+MEI\_Merchandise\_Exports+MEI\_Merchandise\_Imports+MEI\_Nominal\_Retail\_Sales+MEI\_Private\_Consumption+MEI\_Real\_Retail\_Sales+MEI\_WPI\_Inflation+MEI\_Price\_Index+MEI\_Trade\_GDP\_Ratio+MEI\_Merchandise\_Trade\_GDP\_Ratio+MEI\_Real\_Nominal\_Sales,ntrain,na.action = na.exclude)

sm<-summary(model)

new\_test<-data.frame(ntest[c(12:32)])

pred<-predict(model,newdata=new\_test)

df1<-data.frame(UID=u[i,],Pred\_19Q2=pred[[1]],Pred\_19Q3=pred[[2]])

df<-rbind(df,df1)

i=i+1

}

df