Subset data for each source:

googlesearch <- subset(P5\_train1,P5\_train1$Source == "GoogleSearch")

directlink <- subset(P5\_train1,P5\_train1$Source == "Directlink")

googlead <- subset(P5\_train1,P5\_train1$Source == "GoogleAdWord")

Subset data for google:

googles.1 <- subset(googlesearch, googlesearch$Activity == 1)

googles.2 <- subset(googlesearch, googlesearch$Activity == 2)

googles.3 <- subset(googlesearch, googlesearch$Activity == 3)

googles.4 <- subset(googlesearch, googlesearch$Activity == 4)

googles.5 <- subset(googlesearch, googlesearch$Activity == 5)

googles.6 <- subset(googlesearch, googlesearch$Activity == 6)

googles.7 <- subset(googlesearch, googlesearch$Activity == 7)

googles.8 <- subset(googlesearch, googlesearch$Activity == 8)

googles.9 <- subset(googlesearch, googlesearch$Activity == 9)

googles.10 <- subset(googlesearch, googlesearch$Activity == 10)

googles.11 <- subset(googlesearch, googlesearch$Activity == 11)

googles.12 <- subset(googlesearch, googlesearch$Activity == 12)

googles.13 <- subset(googlesearch, googlesearch$Activity == 13)

googles.14 <- subset(googlesearch, googlesearch$Activity == 14)

googles.15 <- subset(googlesearch, googlesearch$Activity == 15)

googles.16 <- subset(googlesearch, googlesearch$Activity == 16)

googles.17 <- subset(googlesearch, googlesearch$Activity == 17)

googles.18 <- subset(googlesearch, googlesearch$Activity == 18)

googles.19 <- subset(googlesearch, googlesearch$Activity == 19)

googles.20 <- subset(googlesearch, googlesearch$Activity == 20)

googles.21 <- subset(googlesearch, googlesearch$Activity == 21)

(trees)

tree.googles.1 <- lm(Revenue ~ Duration, data = googles.1)

tree.googles.2 <- lm(Revenue ~ Duration, data = googles.2)

tree.googles.3 <- lm(Revenue ~ Duration, data = googles.3)

tree.googles.4 <- lm(Revenue ~ Duration, data = googles.4)

tree.googles.5 <- lm(Revenue ~ Duration, data = googles.5)

tree.googles.6 <- lm(Revenue ~ Duration, data = googles.6)

tree.googles.7 <- lm(Revenue ~ Duration, data = googles.7)

tree.googles.8 <- lm(Revenue ~ Duration, data = googles.8)

tree.googles.9 <- lm(Revenue ~ Duration, data = googles.9)

tree.googles.10 <- lm(Revenue ~ Duration, data = googles.10)

tree.googles.11 <- lm(Revenue ~ Duration, data = googles.11)

tree.googles.12 <- lm(Revenue ~ Duration, data = googles.12)

tree.googles.13 <- lm(Revenue ~ Duration, data = googles.13)

tree.googles.14 <- lm(Revenue ~ Duration, data = googles.14)

tree.googles.15 <- lm(Revenue ~ Duration, data = googles.15)

tree.googles.16 <- lm(Revenue ~ Duration, data = googles.16)

tree.googles.17 <- lm(Revenue ~ Duration, data = googles.17)

tree.googles.18 <- lm(Revenue ~ Duration, data = googles.18)

tree.googles.19 <- lm(Revenue ~ Duration, data = googles.19)

tree.googles.20 <- lm(Revenue ~ Duration, data = googles.20)

tree.googles.21 <- lm(Revenue ~ Duration, data = googles.21)

(predictors)

p <- predict(tree.googles.1, newdata = googles.1)

googles.1$PREDICTED <- p

p <- predict(tree.googles.2, newdata = googles.2)

googles.2$PREDICTED <- p

p <- predict(tree.googles.3, newdata = googles.3)

googles.3$PREDICTED <- p

p <- predict(tree.googles.4, newdata = googles.4)

googles.4$PREDICTED <- p

p <- predict(tree.googles.5, newdata = googles.5)

googles.5$PREDICTED <- p

p <- predict(tree.googles.6, newdata = googles.6)

googles.6$PREDICTED <- p

p <- predict(tree.googles.7, newdata = googles.7)

googles.7$PREDICTED <- p

p <- predict(tree.googles.8, newdata = googles.8)

googles.8$PREDICTED <- p

p <- predict(tree.googles.9, newdata = googles.9)

googles.9$PREDICTED <- p

p <- predict(tree.googles.10, newdata = googles.10)

googles.10$PREDICTED <- 10

p <- predict(tree.googles.11, newdata = googles.11)

googles.11$PREDICTED <- p

p <- predict(tree.googles.12, newdata = googles.12)

googles.12$PREDICTED <- p

p <- predict(tree.googles.13, newdata = googles.13)

googles.13$PREDICTED <- p

p <- predict(tree.googles.14, newdata = googles.14)

googles.14$PREDICTED <- p

p <- predict(tree.googles.15, newdata = googles.15)

googles.15$PREDICTED <- p

p <- predict(tree.googles.16, newdata = googles.16)

googles.16$PREDICTED <- p

p <- predict(tree.googles.17, newdata = googles.17)

googles.17$PREDICTED <- p

p <- predict(tree.googles.18, newdata = googles.18)

googles.18$PREDICTED <- p

p <- predict(tree.googles.19, newdata = googles.19)

googles.19$PREDICTED <- p

p <- predict(tree.googles.20, newdata = googles.20)

googles.20$PREDICTED <- p

p <- predict(tree.googles.21, newdata = googles.21)

googles.21$PREDICTED <- p

For < 0’s:

googles.1$PREDICTED[googles.1$PREDICTED < 8] = 0

googles.2$PREDICTED[googles.2$PREDICTED < 3.6] = 0

googles.3$PREDICTED[googles.3$PREDICTED < 0] = 0

googles.4$PREDICTED[googles.4$PREDICTED < 0] = 0

googles.5$PREDICTED[googles.5$PREDICTED < 0] = 0

googles.6$PREDICTED[googles.6$PREDICTED < 0] = 0

googles.7$PREDICTED[googles.7$PREDICTED < 0] = 0

googles.8$PREDICTED[googles.8$PREDICTED < 0] = 0

googles.9$PREDICTED[googles.9$PREDICTED < 0] = 0

googles.10$PREDICTED[googles.10$PREDICTED < 0] = 0

googles.11$PREDICTED[googles.11$PREDICTED < 0] = 0

googles.12$PREDICTED[googles.12$PREDICTED < 0] = 0

googles.13$PREDICTED[googles.13$PREDICTED < 0] = 0

googles.14$PREDICTED[googles.14$PREDICTED < 0] = 0

googles.15$PREDICTED[googles.15$PREDICTED < 0] = 0

googles.16$PREDICTED[googles.16$PREDICTED < 0] = 0

googles.17$PREDICTED[googles.17$PREDICTED < 0] = 0

googles.18$PREDICTED[googles.18$PREDICTED < 0] = 0

googles.19$PREDICTED[googles.19$PREDICTED < 0] = 0

Binding...

final.gs <- rbind(googles.1,googles.2,googles.3,googles.4,googles.5,googles.6,googles.7,googles.8,googles.9,googles.10,googles.11,googles.12,googles.13,googles.14,googles.15,googles.16,googles.17,googles.18,googles.19,googles.20,googles.21)

mean((final.gs$PREDICTED-final.gs$Revenue)^2)

[1] 0.8396814

Actual Submission:

googlesearch.test <- subset(P5\_test\_students,P5\_test\_students$Source == "GoogleSearch")

directlink.test <- subset(P5\_test\_students,P5\_test\_students$Source == "Directlink")

googlead.test <- subset(P5\_test\_students,P5\_test\_students$Source == "GoogleAdWord")

googles.1.test <- subset(googlesearch.test, googlesearch.test$Activity == 1)

googles.2.test <- subset(googlesearch.test, googlesearch.test$Activity == 2)

googles.3.test <- subset(googlesearch.test, googlesearch.test$Activity == 3)

googles.4.test <- subset(googlesearch.test, googlesearch.test$Activity == 4)

googles.5.test <- subset(googlesearch.test, googlesearch.test$Activity == 5)

googles.6.test <- subset(googlesearch.test, googlesearch.test$Activity == 6)

googles.7.test <- subset(googlesearch.test, googlesearch.test$Activity == 7)

googles.8.test <- subset(googlesearch.test, googlesearch.test$Activity == 8)

googles.9.test <- subset(googlesearch.test, googlesearch.test$Activity == 9)

googles.10.test <- subset(googlesearch.test, googlesearch.test$Activity == 10)

googles.11.test <- subset(googlesearch.test, googlesearch.test$Activity == 11)

googles.12.test <- subset(googlesearch.test, googlesearch.test$Activity == 12)

googles.13.test <- subset(googlesearch.test, googlesearch.test$Activity == 13)

googles.14.test <- subset(googlesearch.test, googlesearch.test$Activity == 14)

googles.15.test <- subset(googlesearch.test, googlesearch.test$Activity == 15)

googles.16.test <- subset(googlesearch.test, googlesearch.test$Activity == 16)

googles.17.test <- subset(googlesearch.test, googlesearch.test$Activity == 17)

googles.18.test <- subset(googlesearch.test, googlesearch.test$Activity == 18)

googles.19.test <- subset(googlesearch.test, googlesearch.test$Activity == 19)

googles.20.test <- subset(googlesearch.test, googlesearch.test$Activity == 20)

googles.21.test <- subset(googlesearch.test, googlesearch.test$Activity == 21)

p <- predict(tree.googles.1, newdata = googles.1.test)

googles.1.test$PREDICTED <- p

p <- predict(tree.googles.2, newdata = googles.2.test)

googles.2.test$PREDICTED <- p

p <- predict(tree.googles.3, newdata = googles.3.test)

googles.3.test$PREDICTED <- p

p <- predict(tree.googles.4, newdata = googles.4.test)

googles.4.test$PREDICTED <- p

p <- predict(tree.googles.5, newdata = googles.5.test)

googles.5.test$PREDICTED <- p

p <- predict(tree.googles.6, newdata = googles.6.test)

googles.6.test$PREDICTED <- p

p <- predict(tree.googles.7, newdata = googles.7.test)

googles.7.test$PREDICTED <- p

p <- predict(tree.googles.8, newdata = googles.8.test)

googles.8.test$PREDICTED <- p

p <- predict(tree.googles.9, newdata = googles.9.test)

googles.9.test$PREDICTED <- p

p <- predict(tree.googles.10, newdata = googles.10.test)

googles.10.test$PREDICTED <- 10

p <- predict(tree.googles.11, newdata = googles.11.test)

googles.11.test$PREDICTED <- p

p <- predict(tree.googles.12, newdata = googles.12.test)

googles.12.test$PREDICTED <- p

p <- predict(tree.googles.13, newdata = googles.13.test)

googles.13.test$PREDICTED <- p

p <- predict(tree.googles.14, newdata = googles.14.test)

googles.14.test$PREDICTED <- p

p <- predict(tree.googles.15, newdata = googles.15.test)

googles.15.test$PREDICTED <- p

p <- predict(tree.googles.16, newdata = googles.16.test)

googles.16.test$PREDICTED <- p

p <- predict(tree.googles.17, newdata = googles.17.test)

googles.17.test$PREDICTED <- p

p <- predict(tree.googles.18, newdata = googles.18.test)

googles.18.test$PREDICTED <- p

p <- predict(tree.googles.19, newdata = googles.19.test)

googles.19.test$PREDICTED <- p

p <- predict(tree.googles.20, newdata = googles.20.test)

googles.20.test$PREDICTED <- p

p <- predict(tree.googles.21, newdata = googles.21.test)

googles.21.test$PREDICTED <- p

googles.21.test$PREDICTED <- p

googles.1.test$PREDICTED[googles.1.test$PREDICTED < 8] = 0

googles.2.test$PREDICTED[googles.2.test$PREDICTED < 3.6] = 0

googles.3.test$PREDICTED[googles.3.test$PREDICTED < 0] = 0

googles.4.test$PREDICTED[googles.4.test$PREDICTED < 0] = 0

googles.5.test$PREDICTED[googles.5.test$PREDICTED < 0] = 0

googles.6.test$PREDICTED[googles.6.test$PREDICTED < 0] = 0

googles.7.test$PREDICTED[googles.7.test$PREDICTED < 0] = 0

googles.8.test$PREDICTED[googles.8.test$PREDICTED < 0] = 0

googles.9.test$PREDICTED[googles.9.test$PREDICTED < 0] = 0

googles.10.test$PREDICTED[googles.10.test$PREDICTED < 0] = 0

googles.11.test$PREDICTED[googles.11.test$PREDICTED < 0] = 0

googles.12.test$PREDICTED[googles.12.test$PREDICTED < 0] = 0

googles.13.test$PREDICTED[googles.13.test$PREDICTED < 0] = 0

googles.14.test$PREDICTED[googles.14.test$PREDICTED < 0] = 0

googles.15.test$PREDICTED[googles.15.test$PREDICTED < 0] = 0

googles.16.test$PREDICTED[googles.16.test$PREDICTED < 0] = 0

googles.17.test$PREDICTED[googles.17.test$PREDICTED < 0] = 0

googles.18.test$PREDICTED[googles.18.test$PREDICTED < 0] = 0

googles.19.test$PREDICTED[googles.19.test$PREDICTED < 0] = 0

Final.test <- rbind(googles.1.test,googles.2.test,googles.3.test,googles.4.test,googles.5.test,googles.6.test,googles.7.test,googles.8.test,googles.9.test,googles.10.test,googles.11.test,googles.12.test,googles.13.test,googles.14.test,googles.15.test,googles.16.test,googles.17.test,googles.18.test,googles.19.test,googles.20.test,googles.21.test)

lm <- lm(Revenue ~ Activity + Duration, data = googlead)

decision <- predict(lm, newdata = googlead.test)

googlead.test$PREDICTED <- decision

googlead.test$PREDICTED[googlead$PREDICTED < .08] = 0

googlead.test$PREDICTED[googlead.test$PREDICTED < .08] = 0

lm <- lm(Revenue ~ Activity + Duration, data = directlink)

decision <- predict(lm, newdata = directlink.test)

directlink.test$PREDICTED <- decision

Final.sub <- rbind(Final.test,directlink.test,googlead.test)