report3_575

2023-10-30

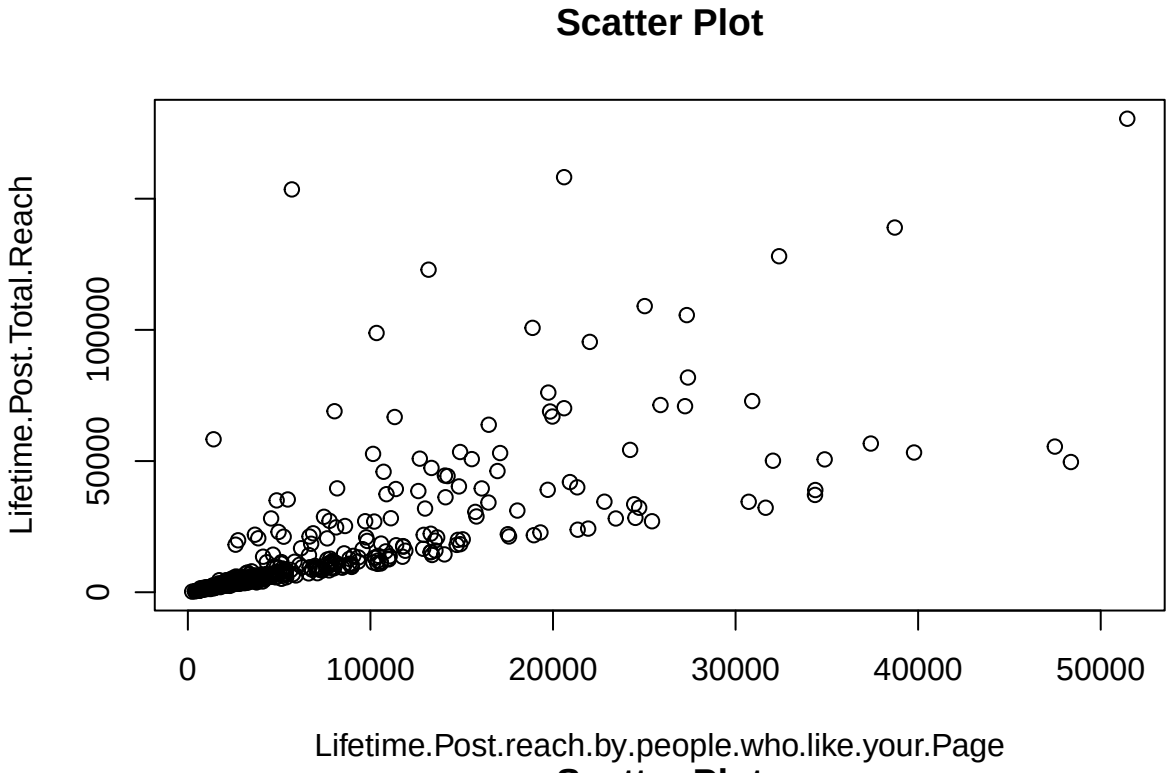
library(GGally) library(ggplot2) library(dplyr) library(car) library(readr) library(MASS)

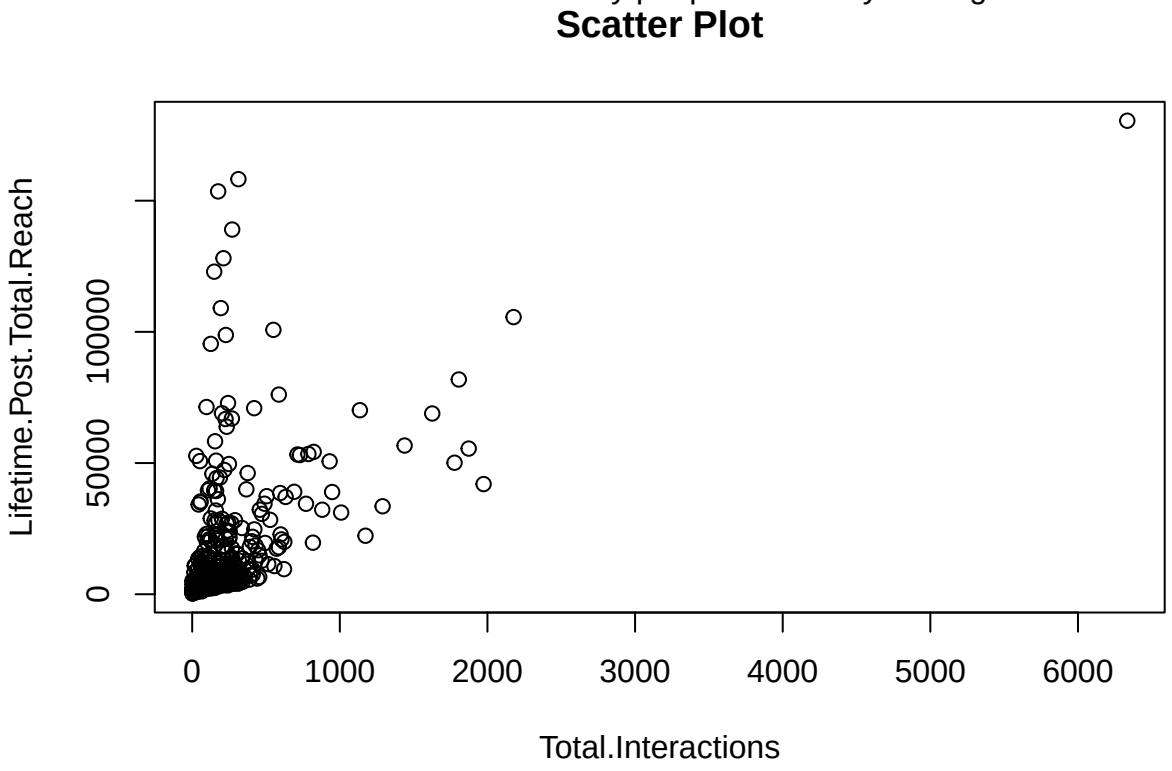
Data loading and preprocessing

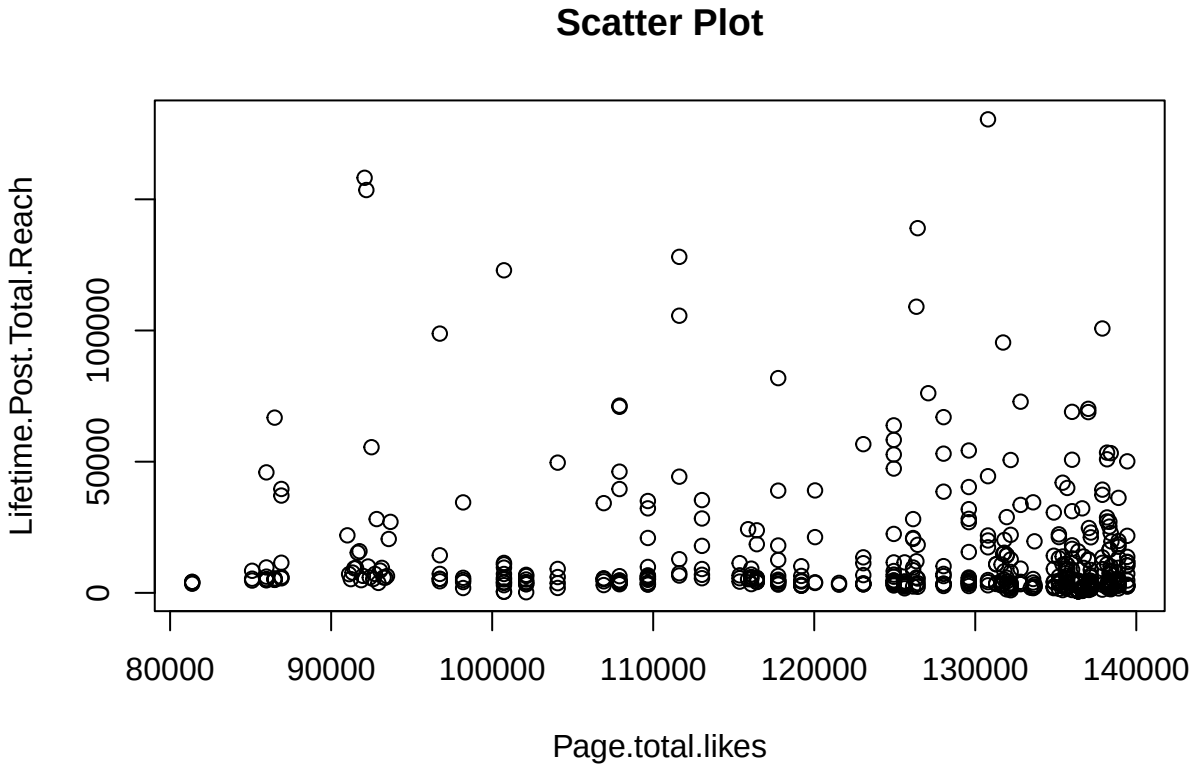
For the 'Paid' column, replace missing values with the mode For 'like' and 'share', replace missing values with their medians

```
data <- read.csv(file="../fb.csv",header = T,sep = ";")
mode_paid <- as.numeric(names(sort(table(data$Paid), decreasing=TRUE)[1]))
data$Paid[is.na(data$Paid)] <- mode_paid

data$Lifetime.Post.Total.Reach[is.na(data$Lifetime.Post.Total.Reach)] <- median(data$Lifetime.Post.Total

data$Lifetime.Post.reach.by.people.who.like.your.Page[is.na(data$Lifetime.Post.reach.by.people.who.like

data$Total.Interactions[is.na(data$Total.Interactions)] <- median(data$Total.Interactions, na.rm=TRUE)

data$Page.total.likes[is.na(data$Page.total.likes)] <- median(data$Page.total.likes, na.rm=TRUE)
```

Data visualization





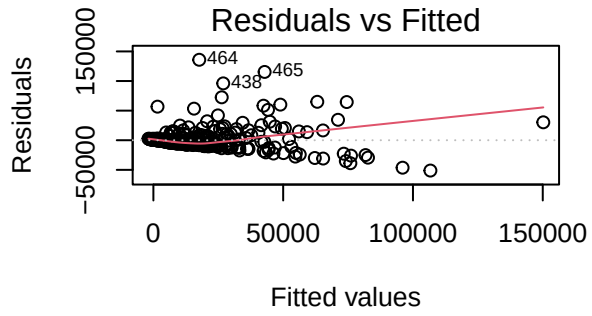


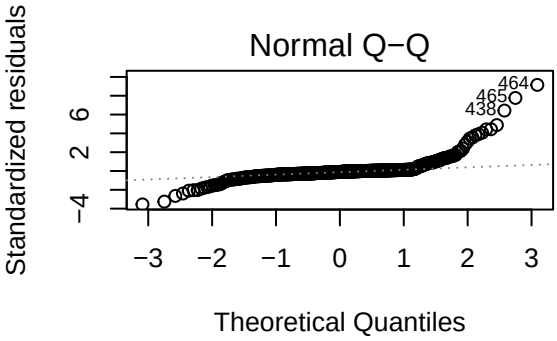
Inital MLR Model

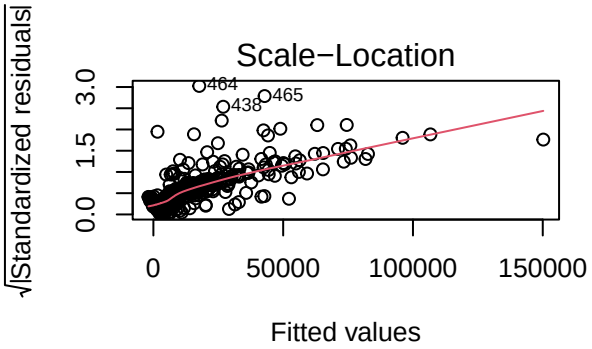
```
##
## Call:
## lm(formula = Lifetime.Post.Total.Reach ~ Paid + Lifetime.Post.reach.by.people.who.like.your.Page +
##     Total.Interactions + Page.total.likes + Post.Month + Post.Weekday +
##     Post.Hour, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -51203  -4521  -1477    681 135869
##
## Coefficients:
##                                                Estimate Std. Error t value
## (Intercept)                                   1.097e+04  1.142e+04   0.961
## Paid                                          3.008e+03  1.519e+03   1.980
## Lifetime.Post.reach.by.people.who.like.your.Page  1.942e+00  1.129e-01  17.202
## Total.Interactions                            7.743e+00  2.272e+00   3.408
## Page.total.likes                             -9.922e-02  1.236e-01  -0.803
## Post.Month                                    1.219e+02  6.124e+02   0.199
## Post.Weekday                                  7.011e+01  3.335e+02   0.210
## Post.Hour                                    -1.604e+02  1.579e+02  -1.016
##                                               Pr(>|t|)
## (Intercept)                                   0.337088
## Paid                                          0.048256 *
## Lifetime.Post.reach.by.people.who.like.your.Page  < 2e-16 ***
## Total.Interactions                            0.000709 ***
## Page.total.likes                              0.422599
## Post.Month                                    0.842319
```

```
## Post.Weekday                                        0.833602
## Post.Hour                                           0.310337
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15030 on 492 degrees of freedom
## Multiple R-squared:  0.5695, Adjusted R-squared:  0.5633
## F-statistic: 92.97 on 7 and 492 DF,  p-value: < 2.2e-16
```

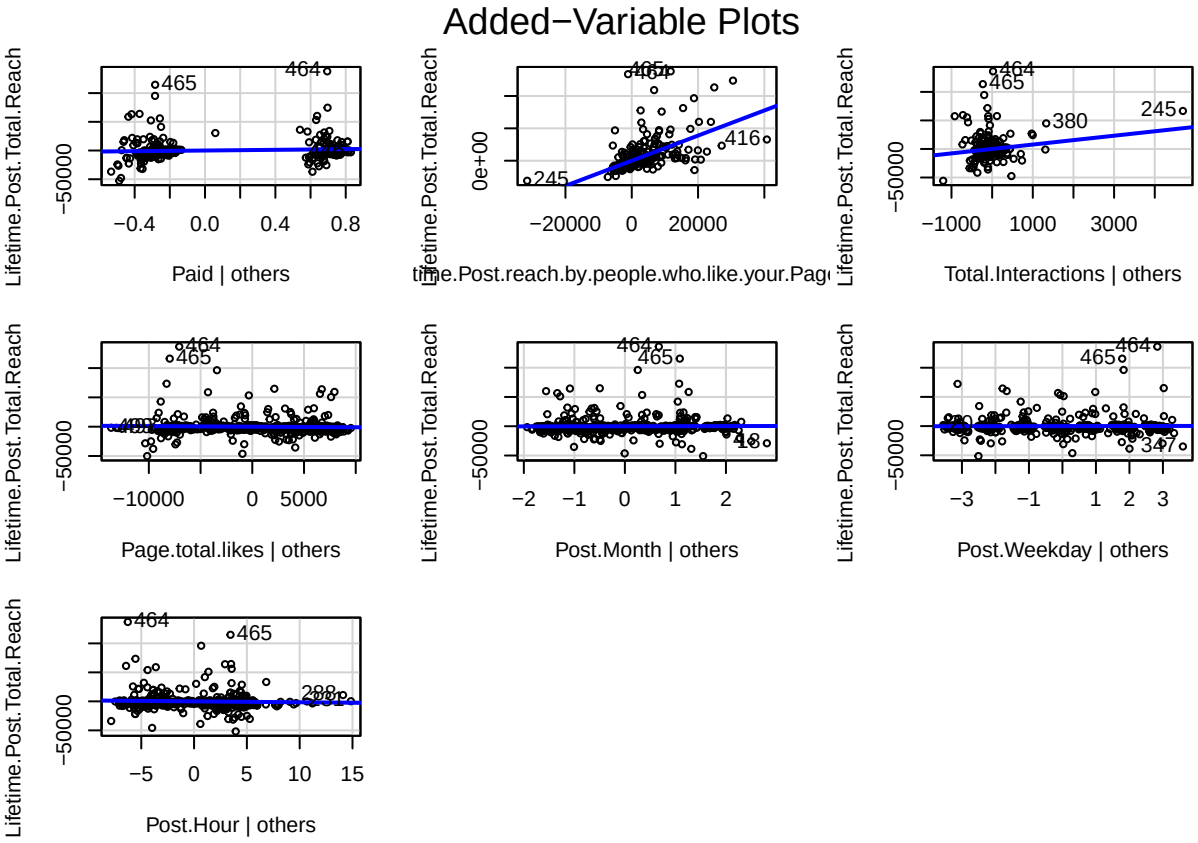
Inital MLR Model Diagnostics







Inital MLR Model AV Plots

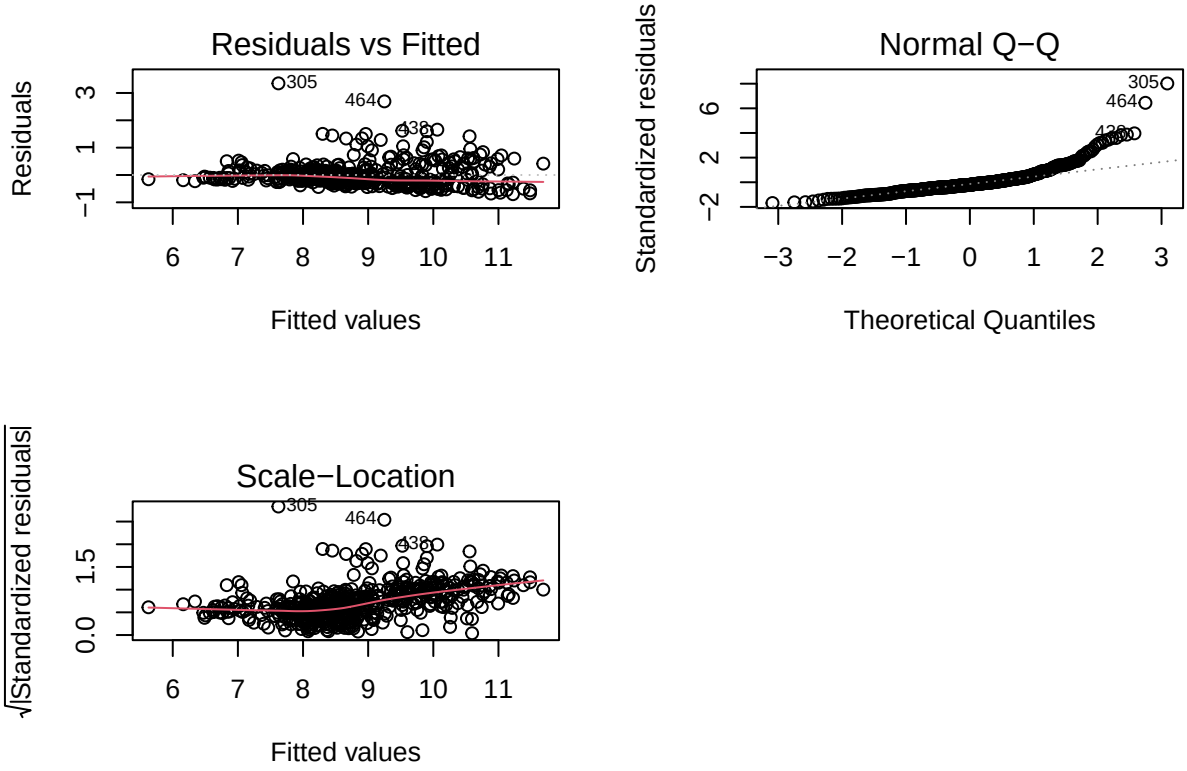


Log-log transformed Model (fm)

```
##
## Call:
## lm(formula = log(Lifetime.Post.Total.Reach) ~ Paid + log(Lifetime.Post.reach.by.people.who.like.your
##     log(Total.Interactions), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7013 -0.2256 -0.0871  0.1091  3.3512
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                       -0.338758   0.168937
## Paid                                               0.101128   0.042425
## log(Lifetime.Post.reach.by.people.who.like.your.Page)  1.090569   0.021340
## log(Total.Interactions)                            0.010799   0.006231
##                                                   t value Pr(>|t|)
## (Intercept)                                        -2.005   0.0455 *
## Paid                                                2.384   0.0175 *
## log(Lifetime.Post.reach.by.people.who.like.your.Page)  51.105   <2e-16 ***
## log(Total.Interactions)                             1.733   0.0837 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.4192 on 496 degrees of freedom
## Multiple R-squared:  0.8662, Adjusted R-squared:  0.8654
## F-statistic:  1070 on 3 and 496 DF,  p-value: < 2.2e-16
```

Log-log transformed Model Diagnostics



Handle Bad Leverage Points

hat_values <- hatvalues(fm) std_residuals <- rstandard(fm) p = 3 n = 500 high_leverage <- which(hat_values > (2*(p+1))/n) outliers <- which(abs(std_residuals) > 2)

to_remove <- intersect(high_leverage, outliers) length(to_remove)

```
## [1] 0
```

Log-log transformed Model AV Plots