Facebook Data Analysis

Ahnaf Ryan: 20757532

February 16, 2021

Case Study

This is a case study about the effectiveness of various posts of a cosmetics company facebook's page.

Importing the Data

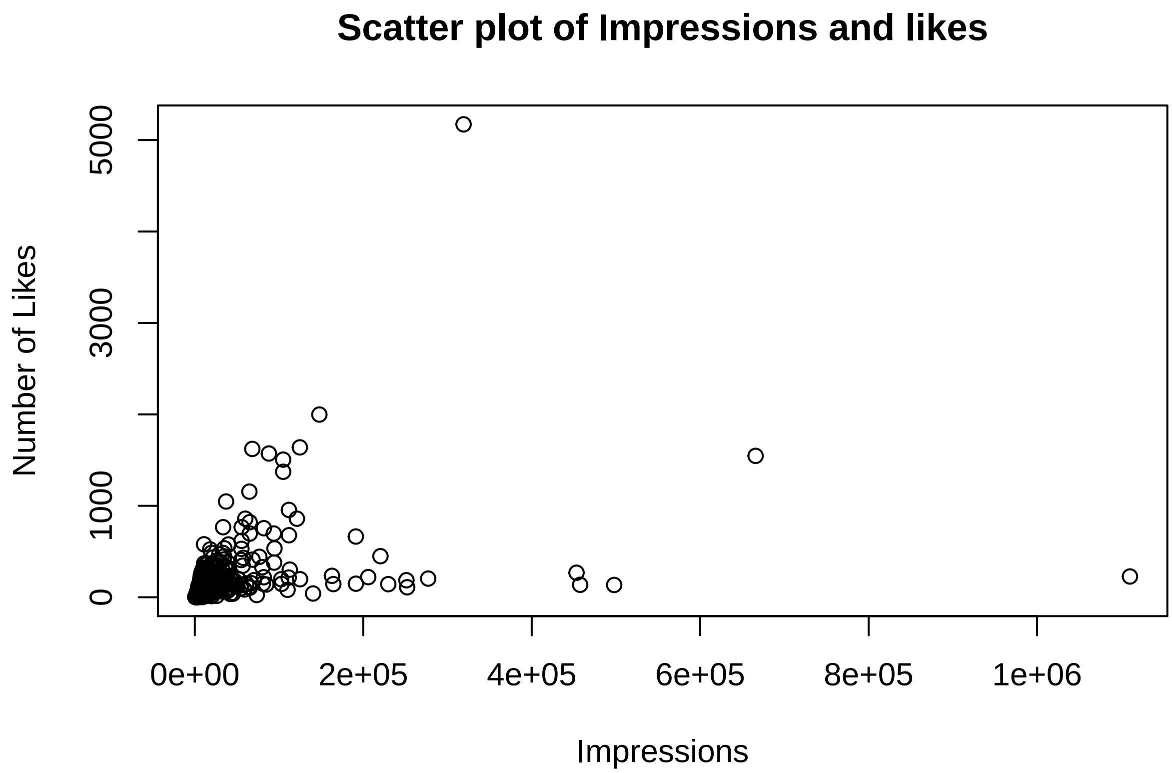
```
fb<-read.csv('facebook.csv')
head(fb)
```

```
##   All.interactions share like comment Impressions.when.page.liked Impressions
## 1              100    17   79       4                        3078        5091
## 2              164    29  130       5                       11710       19057
## 3               80    14   66       0                        2812        4373
## 4             1777   147 1572      58                       61027       87991
## 5              393    49  325      19                        6228       13594
## 6              186    33  152       1                       16034       20849
##   Paid Post.Hour Post.Weekday Post.Month    Category   Type Page.likes
## 1    0         3            4         12     Product  Photo     139441
## 2    0        10            3         12     Product Status     139441
## 3    0         3            3         12 Inspiration  Photo     139441
## 4    1        10            2         12     Product  Photo     139441
## 5    0         3            2         12     Product  Photo     139441
## 6    0         9            1         12     Product Status     139441
```

What is the typical number of likes from a facebook post?

We are interested in the measure of location and the word "typical" may suggest to look for mean , median or mode. However, it is common knowledge to think that more views of a post will likely increase the number of likes, hence we need to weight the post based on the impressions.

```
subdata<-na.omit(fb)
plot(subdata$Impressions,subdata$like,ylab="Number of Likes",xlab="Impressions",
     main="Scatter plot of Impressions and likes")
```



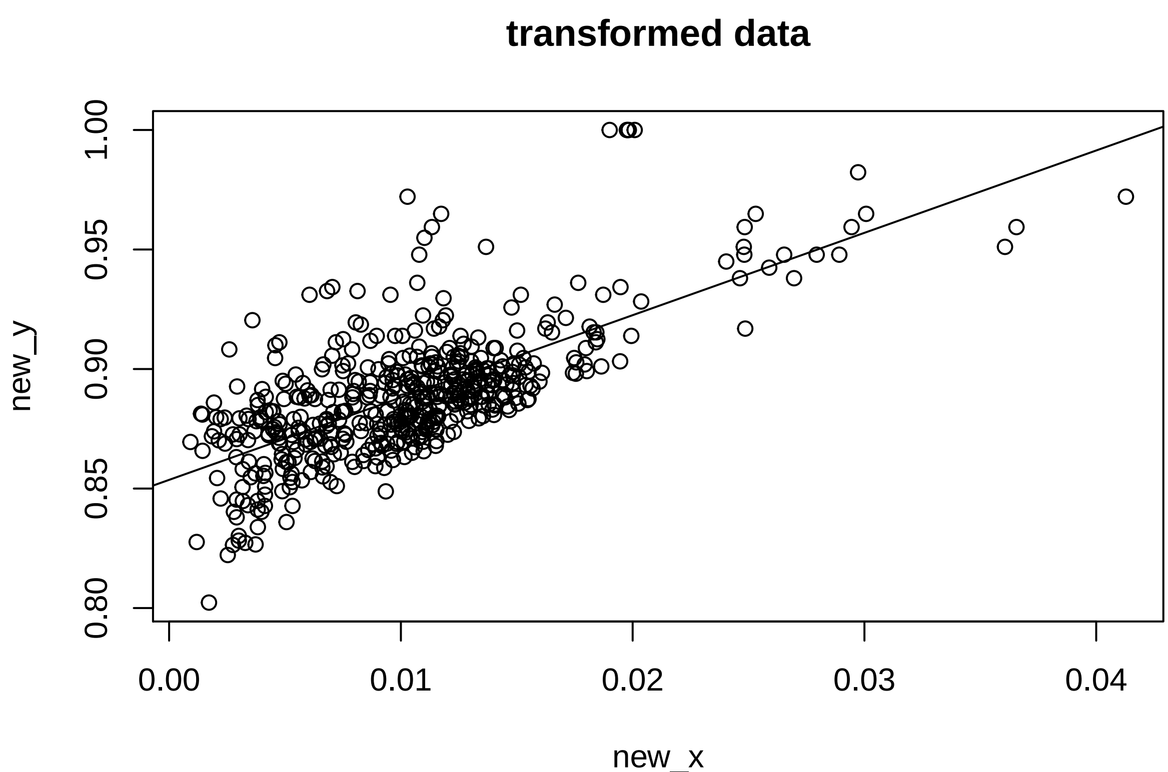
From the scatter plot, it is difficult to understand the relationship between Number of Likes and Impressions. To have a better understanding, let us apply power transformation to the variates.

```
library("MASS")
powerfun<-function(x,alpha) {
  if(sum(x<=0) > 0) stop("x must be positive")
  if (alpha==0)
    log(x)
  else if (alpha>0) {
    x^alpha
  }
  else -x^alpha
}
rho_cor<-function(alpha,x,y) {
  alphax<-alpha[1]
  alphay<-alpha[2]
  return(-abs(cor(x=powerfun(x,alphax),y=powerfun(y,alphay))))
}
min3<-nlminb(start=c(1,1),objective = rho_cor,y=subdata$like+1,x=subdata$Impressions)
print(min3)
```

```
## $par
## [1] -0.50228247 -0.02576016
##
## $objective
## [1] -0.6954489
##
```

```
## $convergence
## [1] 0
##
## $iterations
## [1] 8
##
## $evaluations
## function gradient
##        9       22
##
## $message
## [1] "relative convergence (4)"
```

```
new_y<-(subdata$like +1)^(min3$par[2])
new_x<-(subdata$Impressions)^(min3$par[1])
plot(new_x,new_y,main="transformed data")
abline(lm(new_y~new_x))
```



We optimize the correlation function and choose the best power parameters of our variates. After applying the transformation, we see they have a expotential relationship which is a interesting observation because we can confirm that there is a relation between number of likes and expressions and we know their structure based on the dataset.

To find the "typical" number of likes, we use a robust mean estimator known "Tukey's Biweight". We see the "typical" number of likes is 111 likes per post.

```
library("DescTools")
```

```
## Warning: package 'DescTools' was built under R version 4.0.3
```

```r
TukeyBiweight(subdata$like)
```

```
## [1] 111.5795
```

We want to test the hypothesis whether the investment made into facebook advertisement has impacted the number of likes of a post for a particular category. The reason to test this hypothesis is that if there is not statistical significance between paid and unpaid, then we can redistribute our investment into categories that are affected or statistically significant.From the aggregate, it seems like category "inspiration", the values did not change significantly. For the category "inspiration", We define $H_o = \mu_(unpaid) = \mu_(paid)$.We find that the p-value is 0.21466556 hence there is no evidence against the null hypothesis based on the observed data. These proves that based on the observed,the investment for the category "inspiration" has no effect on the number of likes. Hence we allocate this investment into the other categories for effective investment.

```r
head(subdata)
```

```
##   All.interactions share like comment Impressions.when.page.liked Impressions
## 1              100    17   79       4                        3078        5091
## 2              164    29  130       5                       11710       19057
## 3               80    14   66       0                        2812        4373
## 4             1777   147 1572      58                       61027       87991
## 5              393    49  325      19                        6228       13594
## 6              186    33  152       1                       16034       20849
##   Paid Post.Hour Post.Weekday Post.Month    Category   Type Page.likes
## 1    0         3            4         12     Product  Photo     139441
## 2    0        10            3         12     Product Status     139441
## 3    0         3            3         12 Inspiration  Photo     139441
## 4    1        10            2         12     Product  Photo     139441
## 5    0         3            2         12     Product  Photo     139441
## 6    0         9            1         12     Product Status     139441
```

```r
paid_data<-subdata[subdata$Paid==1,]
unpaid_data<-subdata[subdata$Paid==0,]
TukeyBiweight(paid_data$like)
```

```
## [1] 133.6484
```

```r
setNames(aggregate(x=paid_data$like,by=list(paid_data$Category),FUN=mean),c("Categories","Likes"))
```

```
##    Categories    Likes
## 1      Action 151.0625
## 2 Inspiration 221.6512
## 3     Product 423.6250
```

```r
setNames(aggregate(x=unpaid_data$like,by=list(unpaid_data$Category),FUN=mean),c("Categories","Likes"))
```

```
##    Categories    Likes
## 1      Action 115.8571
## 2 Inspiration 215.2321
## 3     Product 152.4227
```

```r
proportion_paid<-length(which(paid_data$Category=="Inspiration"))/139
proportion_unpaid<-length(which(unpaid_data$Category=="Inspiration"))/495
ins_paid_data<-paid_data[paid_data$Category=="Inspiration",]
ins_unpaid_data<-unpaid_data[unpaid_data$Category=="Inspiration",]
v1<-(1/(nrow(ins_unpaid_data)-1))*var(ins_unpaid_data$like)
v2<-(1/(nrow(ins_paid_data)-1))*var(ins_paid_data$like)
d=(mean(ins_paid_data$like)-mean(ins_unpaid_data$like))/sqrt((v1/nrow(ins_unpaid_data))+(v2/nrow(ins_pai
1-pt(d,df=nrow(ins_paid_data)+nrow(ins_unpaid_data)-2)
```

```
## [1] 0.2146556
```

We can see 30 percent of the paid advertisements were directed to inspirational categories which seems a waste of advertisement money in terms of number of likes.

```r
aggregate(x=paid_data$Impressions,by=list(paid_data$Category),FUN=mean)
```

```
##       Group.1        x
## 1      Action 47675.92
## 2 Inspiration 26387.09
## 3     Product 31485.78
```

```r
aggregate(x=unpaid_data$Impressions,by=list(unpaid_data$Category),FUN=mean)
```

```
##       Group.1        x
## 1      Action 38280.71
## 2 Inspiration 24030.20
## 3     Product 13062.99
```

```r
length(which(paid_data$Category=="Inspiration"))/139
```

```
## [1] 0.3093525
```

```r
length(which(unpaid_data$Category=="Inspiration"))/495
```

```
## [1] 0.2262626
```

Let us now perform the hypothesis whether the investment made into facebook advertisement has impacted the impression for a particular category.

```r
setNames(aggregate(x=paid_data$like,by=list(paid_data$Type),FUN=mean),c("Categories","Likes"))
```

```
##   Categories     Likes
## 1       Link  56.66667
## 2      Photo 240.88235
## 3     Status 277.80000
## 4      Video 243.00000
```

```r
setNames(aggregate(x=unpaid_data$like,by=list(unpaid_data$Type),FUN=mean),c("Categories","Likes"))
```

```
##   Categories    Likes
## 1       Link  79.5625
## 2      Photo 161.6788
## 3     Status 147.8286
## 4      Video 216.0000
```

```r
length(which(paid_data$Type=="Link"))/139
```

```
## [1] 0.04316547
```

```r
length(which(unpaid_data$Type=="Link"))/495
```

```
## [1] 0.03232323
```