BY: Ivan Carrasco , Jeff Lin, Dhanahsree Kharate , Olumuyide Omole

EXECUTIVE SUMMARY

This report provides an in-depth analysis on how participation and elaboration affect customer purchasing habits when introducing a loyalty program contest (Block Party). Do contests increase revenue? In order to examine this, the MSIT conducted a multiple regression model analysis using R, along with Tukey's first aid technique (variance stabilizing transformation). Results of the analysis present an initial increase in customer purchases directly after the contest is introduced, while slowing down as time passes. All calculations are displayed in the analysis section of the report.

Introduction

In order to maximize profits and exposure, sponsor companies nowadays are incorporating more "loyalty" type programs. These programs often allow customers to accumulate points with purchases from parent company products. Typically after accumulating a certain amount of points, customers will exchange the points for better saving deals, discounts, or freebies on various products or services that are offered.

In the Block Party Project assignment, a data set is analyzed from a leading coalition program (LP). With 10 million plus members, the LP covers roughly 67% of families within the given country. In addition, the LP program is sponsored by over 100 parent companies from varying categories: retail, groceries, gas, credit card purchases, etc. If enrolled in the LP program, a member will collect miles(points) when purchasing products or services from one of the 100+ parent company (sponsors). The LP generates revenue each time a mile is acquired by a member. As a result, the fundamental goal for the LP is to have its members accrue as many points as possible.

In order to help achieve this, the LP launched a "Block Party" that promised the general audience (members and nonmembers) 4 opportunities to win 25,000 miles. In order to increase their likelihood of winning, the audience can either become a member, upload photos, or give thumbs up on the community website. One thing to note is that this Block Party allowed members to share personal stories regarding the LP program, as well hear others' stories. As a result, passive audiences were converted to active players.

In order to find the efficacy of the program and Block Party on product/services purchase rates, a study was conducted by the famous MSIT program, along with world renowned Dr.

Malthouse. A dataset of mile accumulation on 143,000 members were given to analyze. With the data, the aim is to assess the following questions:

- To what rate does involvement within the Block party affect purchases? Does it really make a difference?
- Is the number of words a person posts (elaboration) to the social media website directly correlated to purchases?
- What is the length of involvement and elaboration regarding the Block Party?

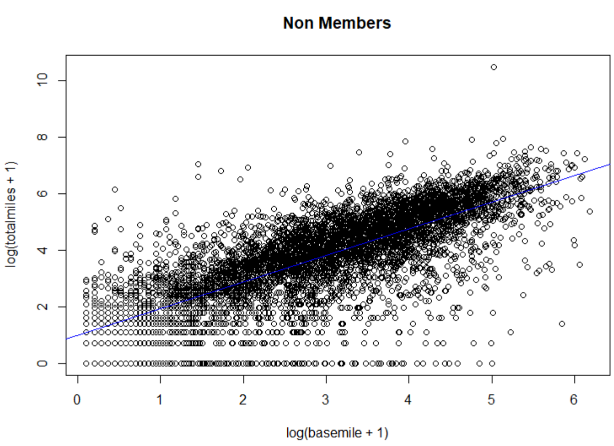
Analysis and Method Section

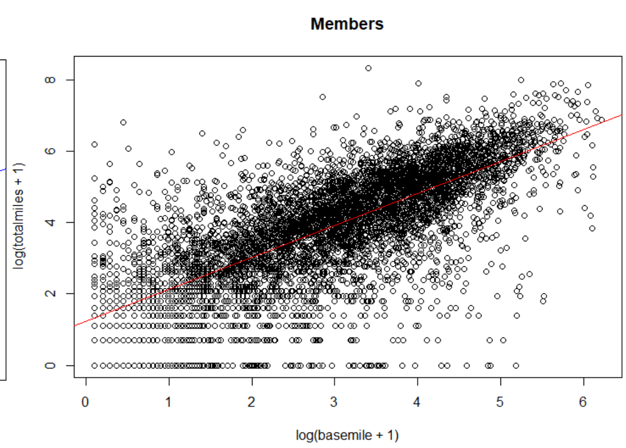
The data consists of number of miles accumulated before the loyalty program in categories such as pregas, prefood, preother, prebank , preretail and number of miles after Block party loyalty program in categories such as mile1,mile2,mile3,mile4.

1. Does participation in the Block Party contest increase subsequent purchases (i.e., mile accumulation)?

In order to answer this question, we started formulating a model for this purpose, with this model we will understand how significant are the variables in our model and how are the influence over the model we construct.

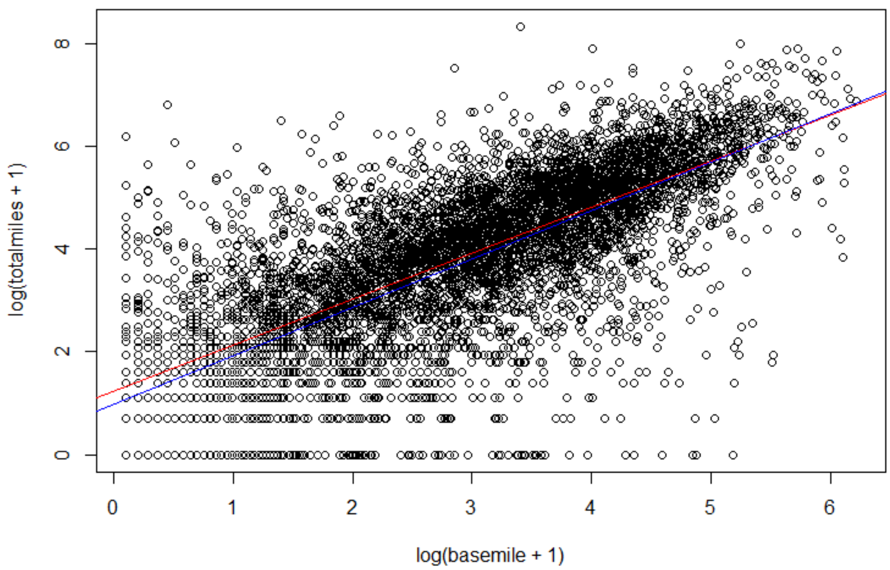
We defined as 'totalmiles' variable as mile1 + mile2 + mile3+ mile4





The above graph shows the influence of being member, which decrease slightly the slope but the area below the regression in slightly bigger and the intercept as well.

BY: Ivan Carrasco , Jeff Lin, Dhanahsree Kharate , Olumuyide Omole



Now, we want to analyze the model we construct and understand if the variable 'particpate' is significant:

ln(totalmiles+1) = ln(basemile+1)+ln(particpate+1)

```
> fit_all = lm(log(totalmiles+1)~log(basemile+1)+log(particpate+1),party)
> summary(fit_all)
Call:
lm(formula = log(totalmiles + 1) ~ log(basemile + 1) + log(particpate +
    1), data = party)
Residuals:
    Min      1Q  Median      3Q     Max
-5.9177 -0.5335  0.1367  0.6604  5.2446
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.046304   0.025342  41.288  < 2e-16 ***
log(basemile + 1)   0.920830   0.007353 125.225  < 2e-16 ***
log(particpate + 1) 0.151366   0.028644   5.284 1.28e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.101 on 12299 degrees of freedom
Multiple R-squared:  0.5609,  Adjusted R-squared:  0.5609
F-statistic:  7857 on 2 and 12299 DF,  p-value: < 2.2e-16
```

	pre-period	post-period	
		mile1	totalmiles
member	*37.924*	*38.47*	128.1
non-member	*37.243*	*32.98*	119.6

Supporting our conclusion about the area below the regressions, we found an increase in purchases being a member.

With this model, from a statistical perspective, we can't reject the coefficient related to the variable 'particpate' - being member or not -. Anticipating our initial guess we concluded that the variable 'particpate' is significant, and *YES*, the degree of affect is soft but still influent.

2. Does the amount of elaboration (number of words written) affect subsequent purchases (mile accumulation)?

To analyze this relationship we took the model we used before, but this time we add the variable elaboration. We used natural log (log in 'r') because when we make a regression the residual standard error is less and, to eliminate any data from non-member, we created a subset including only the member for this analysis.
the model used is:

log(totalmiles+1) = log(elaboration+1) + log(basemile+1)

** Since all data is taking from party where particpate == 1*

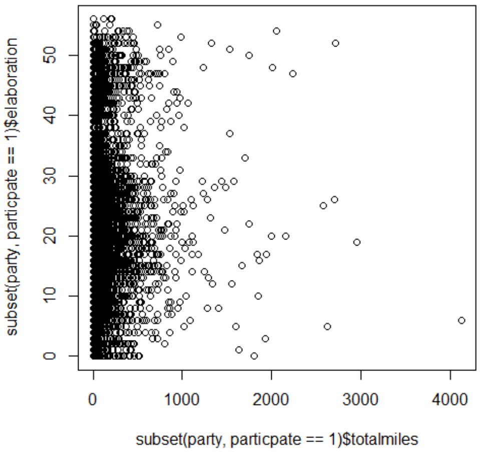
```
> summary(lm(log(totalmiles+1) ~ log(elaboration+1) + log(basemile+1),subset(party,particpate ==
1), main="Elaboration vs Totalmiles"))

Call:
lm(formula = log(totalmiles + 1) ~ log(elaboration + 1) + log(basemile +
    1), data = subset(party, particpate == 1), main = "Elaboration vs Totalmiles")

Residuals:
    Min      1Q  Median      3Q     Max
-5.8806 -0.5690  0.1203  0.6729  5.1786

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           1.16402    0.06493  17.928   <2e-16 ***
log(elaboration + 1)  0.01919    0.01968   0.975     0.33
log(basemile + 1)     0.89685    0.01066  84.094   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.127 on 6148 degrees of freedom
Multiple R-squared:  0.5367,  Adjusted R-squared:  0.5365
F-statistic:  3561 on 2 and 6148 DF,  p-value: < 2.2e-16
```

Since the variable 'elaboration' is not significant from statistical perspective, we can't reject the hypothesis that coefficient can be zero.

```
> summary(lm(log(mile1+1) ~ log(elaboration+1) * log(basemile+1),subset(party,particpate == 1),
main="Elaboration vs Totalmiles"))

Call:
lm(formula = log(mile1 + 1) ~ log(elaboration + 1) * log(basemile +
    1), data = subset(party, particpate == 1), main = "Elaboration vs Totalmiles")

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-4.0442 -1.1545 -0.1411  1.1681  6.1415

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                              0.198422   0.181416   1.094    0.274
log(elaboration + 1)                    -0.003235   0.060813  -0.053    0.958
log(basemile + 1)                        0.581668   0.059740   9.737   <2e-16 ***
log(elaboration + 1):log(basemile + 1)  0.014239   0.019793   0.719    0.472
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.56 on 6147 degrees of freedom
Multiple R-squared:  0.2266,  Adjusted R-squared:  0.2263
F-statistic: 600.5 on 3 and 6147 DF,  p-value: < 2.2e-16
```

Changing the model where '*elaboration'* and '*basemile'* interact together, we got the same results, we couldn't reject the hypothesis because there coefficients are not significant enough.




```
> cor(subset(party,particpate == 1)$totalmiles,subset(party,particpate == 1)$elaboration)
[1] 0.04171225
```

The correlation between 'totalmiles' and 'elabotarion' is close to zero, it means that the variables are not correlated. As a conclusion, the answer is *NO*, the amount of elaboration does not affect subsequent purchases.

3. For how long does the participation / elaboration affect persist?

In order to find the effect of elaboration (number of words) in the posts on the subsequent purchases across the week, we take into consideration the factors such as number of points accumulated per week, basemile (total number of miles before block party) and elaboration factor.

```
> fit3=lm(mile1 ~ elaboration + basemile, subset(party, particpate == 1)) #mile1
> summary(fit3)$coefficients
             Estimate Std. Error   t value      Pr(>|t|)
(Intercept) 4.3636761 2.98808669  1.460358  1.442428e-01
elaboration 0.1302129 0.11200024  1.162612  2.450319e-01
basemile    0.8238948 0.02718168 30.310669 3.304181e-188
> fit3=lm(mile2 ~ elaboration + basemile, subset(party, particpate == 1)) #mile2
> summary(fit3)$coefficients
               Estimate Std. Error     t value      Pr(>|t|)
(Intercept) 8.482310940 2.05992586  4.11777488  3.875612e-05
elaboration 0.001084135 0.07721067  0.01404125  9.887975e-01
basemile    0.594977378 0.01873849 31.75161374 5.166598e-205
> fit3=lm(mile3 ~ elaboration + basemile, subset(party, particpate == 1)) #mile3
> summary(fit3)$coefficients
               Estimate Std. Error   t value      Pr(>|t|)
(Intercept)  9.35406834 2.04376764  4.576875  4.812556e-06
elaboration -0.08538843 0.07660503 -1.114658  2.650404e-01
basemile     0.56010990 0.01859151 30.127193 4.148285e-186
> fit3=lm(mile4 ~ elaboration + basemile, subset(party, particpate == 1)) #mile4
> summary(fit3)$coefficients
             Estimate Std. Error   t value      Pr(>|t|)
(Intercept) 2.7694853 2.54972533  1.086190  2.774377e-01
elaboration 0.1242452 0.09556946  1.300052  1.936320e-01
basemile    0.6407410 0.02319404 27.625239 1.883312e-158
```

Considering the above factors have low multicollinearity, but model has **heteroscedastic error variance having the least square estimates unbiased** thus, we use **logarithm** or square root to variance stabilize the factors to eliminate the error.

Thus, using multiple linear regression using logarithm as variance stabilizing transformation.

```
> fit3=lm(log(mile1+1) ~ log(elaboration+1) + log(basemile+1), subset(party, particpate == 1))
#mile1 log
> summary(fit3)$coefficients
                     Estimate Std. Error    t value  Pr(>|t|)
(Intercept)        0.08506414 0.08989489  0.9462623 0.3440520
log(elaboration + 1) 0.03587568 0.02724844  1.3166141 0.1880171
log(basemile + 1)    0.62331139 0.01476605 42.2124606 0.0000000
> fit3=lm(log(mile2+1) ~ log(elaboration+1) + log(basemile+1), subset(party, particpate == 1))
#mile2 log
> summary(fit3)$coefficients
                        Estimate Std. Error   t value      Pr(>|t|)
(Intercept)           0.29101787 0.08809534  3.303442  9.605156e-04
log(elaboration + 1) -0.02445783 0.02670297 -0.915922  3.597438e-01
log(basemile + 1)     0.57522052 0.01447046 39.751365 9.375490e-308
> fit3=lm(log(mile3+1) ~ log(elaboration+1) + log(basemile+1), subset(party, particpate == 1))
#mile3 log
> summary(fit3)$coefficients
                         Estimate Std. Error    t value      Pr(>|t|)
(Intercept)           0.333810576 0.08606004  3.8788106  1.060643e-04
log(elaboration + 1) -0.007135151 0.02608604 -0.2735237  7.844598e-01
```

```
log(basemile + 1)      0.546731098 0.01413614 38.6761159 4.257952e-293
> fit3=lm(log(mile4+1) ~ log(elaboration+1) + log(basemile+1), subset(party, particpate == 1))
#mile4 log
> summary(fit3)$coefficients
                        Estimate Std. Error    t value      Pr(>|t|)
(Intercept)          -0.07709711 0.08639198 -0.8924105  3.722079e-01
log(elaboration + 1)  0.05319893 0.02618666  2.0315279  4.224439e-02
log(basemile + 1)     0.56370174 0.01419067 39.7234112 2.270265e-307
```

In order to reject the hypothesis Ho: $\beta 1 = 0$ Versus H1: $\beta 1 \neq 0$ we will consider P-value << 0.05 (for 95% confidence interval):

Week	p-value	Reject H_0
1	0.18801	No
2	0.35914	No
3	0.78445	No
4	0.04224	Yes

By observing the above summary, the term elaboration has significant effect on number of miles accumulated in the 4th week.

Testing the significance of elaboration term in the model by null hypothesis,

Ho: $\beta 1 = 0$ Versus H1: $\beta 1 \neq 0$

P = 0.04224 < 0.05 (for 95% confidence interval)

Thus, we can reject the null hypothesis and conclude that elaboration is significant factor affecting the total number of miles for 4th week. Also, observing the slope of elaboration is 0.0531 to compare with the others weeks being the highest.

For total number of miles:

The below is the summary is for total number of miles accumulated regressing against elaboration and basemile. We can see that elaboration has insignificant effect on number of miles post block party.

```
> fit3=lm(log(totalmiles+1) ~ log(elaboration+1) + log(basemile+1), subset(party, particpate ==
1)) #mile4 log
> summary(fit3)

Call:
lm(formula = log(totalmiles + 1) ~ log(elaboration + 1) + log(basemile +
    1), data = subset(party, particpate == 1))

Residuals:
    Min      1Q  Median      3Q     Max
-5.8806 -0.5690  0.1203  0.6729  5.1786

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           1.16402    0.06493  17.928   <2e-16 ***
log(elaboration + 1)  0.01919    0.01968   0.975     0.33
```

```
log(basemile + 1)      0.89685    0.01066  84.094   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.127 on 6148 degrees of freedom
Multiple R-squared:  0.5367,  Adjusted R-squared:  0.5365
F-statistic:  3561 on 2 and 6148 DF,  p-value: < 2.2e-16
```

For studying the effect of participation for number of miles accumulation,

```
> fit3=lm(log(mile1+1) ~ log(particpate+1) + log(basemile+1), subset(party)) #mile1 log
> summary(fit3)$coefficients
                      Estimate Std. Error   t value    Pr(>|t|)
(Intercept)         0.09755344 0.03546662  2.750571 0.005957857
log(particpate + 1) 0.10691645 0.04008852  2.667009 0.007662969
log(basemile + 1)   0.62973741 0.01029139 61.190697 0.000000000
> fit3=lm(log(mile2+1) ~ log(particpate+1) + log(basemile+1), subset(party)) #mile2 log
> summary(fit3)$coefficients
                     Estimate Std. Error   t value     Pr(>|t|)
(Intercept)         0.1539710 0.03480679  4.423590 9.790824e-06
log(particpate + 1) 0.1112258 0.03934271  2.827101 4.704730e-03
log(basemile + 1)   0.5711572 0.01009993 56.550615 0.000000e+00
> fit3=lm(log(mile3+1) ~ log(particpate+1) + log(basemile+1), subset(party)) #mile3 log
> summary(fit3)$coefficients
                      Estimate  Std. Error   t value     Pr(>|t|)
(Intercept)         0.22392152 0.033997107  6.586487 4.686902e-11
log(particpate + 1) 0.09309538 0.038427503  2.422624 1.542323e-02
log(basemile + 1)   0.55523171 0.009864981 56.283098 0.000000e+00
> fit3=lm(log(mile4+1) ~ log(particpate+1) + log(basemile+1), subset(party)) #mile4 log
> summary(fit3)$coefficients
                      Estimate  Std. Error   t value   Pr(>|t|)
(Intercept)         0.07015604 0.033806035  2.075252 0.03798393
log(particpate + 1) 0.05331037 0.038211531  1.395138 0.16299931
log(basemile + 1)   0.55384118 0.009809538 56.459459 0.00000000
```

In order to reject the hypothesis Ho: $\beta 1 = 0$ Versus H1: $\beta 1 \neq 0$ we will consider P-value << 0.05 (for 95% confidence interval):

Week	p-value	Reject H_0
1	0.00766	Yes
2	0.00470	Yes
3	0.01542	Yes
4	0.16299	No

Thus, we reject the hypothesis in all weeks except 4th week.
From the below summary, we can infer there is significant effect on miles accumulated in relation to block party participation in the first 3 weeks (week1, week2, week3), losing the effect in week 4.

The effect reduces week after week and on last week 4 there is no effect of participation on the number of miles accumulated. Thus, the participation effect persists till week 3.

Conclusion

The study was set out to analyze the effectiveness of the Block Party by LP, and whether or not there is a direct correlation between it and customer mile accumulation. This is very important because as a company we need to know what programs or events can be created in order to increase revenue. Is implementing the block party worth it? Should the LP continue to provide these types of events? With this experiment, we determined the above questions by finding whether or not both block party participation and elaboration increased subsequent purchases.

With this in mind, we took the basemiles (miles accumulated pre party) and sum of all miles post party and incorporated 2 factors: participation and elaboration. We performed a regression analysis model across each week in order to find the effectiveness of that week after applying the 2 factors. We have concluded that participation does indeed have an effect for subsequent purchases post block party. The effect of participation can be seen in week 1-3(constant reduction) while no effect is seen in week 4. In other words, as time moves away from block party, members purchases less and less(week 1-3) until there is no correlation (week 4). Similarly, the effect of elaboration also has a constant increment for subsequent purchases overtime. The key difference between the 2 factors is that we see a continuation of increment into week 4 with elaboration compared to participation where the continued reduction stops at week 3.

As a result, we believe it is worth hosting these types of events in the future as it increases subsequent purchases compared to not hosting the event.