Loyalty Program
Block Party Contest

Prepared for:
Dr Edward Malthouse

Prepared by:
Ruma Anand
Akshay Jain
Ademola Popoola
Liping Zhou

May 5, 2018

Executive Summary

After a thorough analysis of the sample Loyalty Program data containing member purchasing and activity behavior resulting from the Block Party contest, it is concluded that, *1) Participation in the Black Party contest had the subsequent effect of increasing purchases by members, thereby accumulating miles. 2) The more elaboration (number of words written) members submit, leads to a subsequent increase in mile accumulation via purchases. 3) There was a visible uptick in subsequent week purchases however, the effect seemingly dissipated following week 3.*

Introduction

The Loyalty Program enjoys a membership of 10 million customers representing over 67% of the households in the country in which it operates. These members accumulate miles after completing purchases from over 100 sponsors. The sponsors offer retail items for sale ranging from groceries, to gasoline to apparel and credit card products. Members can in turn opt to redeem the accumulated miles for rewards such as travel, merchandise, and discount coupons. The Loyalty Program receives compensation from sponsors when a mile is redeemed by a LP member.

In order for the Loyalty Program to generate as much revenue as possible, it must continuously determine the best means for motivating members to make purchases at sponsoring vendors; in the process accumulating and redeeming miles. If the miles are not redeemed, the LP is not receiving revenue from sponsors.

In an effort to generate enthusiasm for the program, the LP launched a Block Party contest. The contest presented opportunities for members to win 25,000 miles through the completion of specific tasks. It also allowed for the LP to observe the effect of the contest and its requirements on purchasing behavior.

The purpose of this report is to analyse the behavior data associated with the contest and present findings answering the following questions:

1. Does participation in the Block Party contest increase subsequent purchases?
2. Does the amount of elaboration affect subsequent purchases?
3. For how long does the participation/elaboration affect persist?

Analysis and Methods

This section begins with a basic analysis of the membership sample data. We then continue to investigate attribution of the change in behavior, post Block Party, through the use of linear regression and multiple linear regression models.

As first step towards basic analysis we looked for the outliers and correlations between independent variables (columns) and found following:

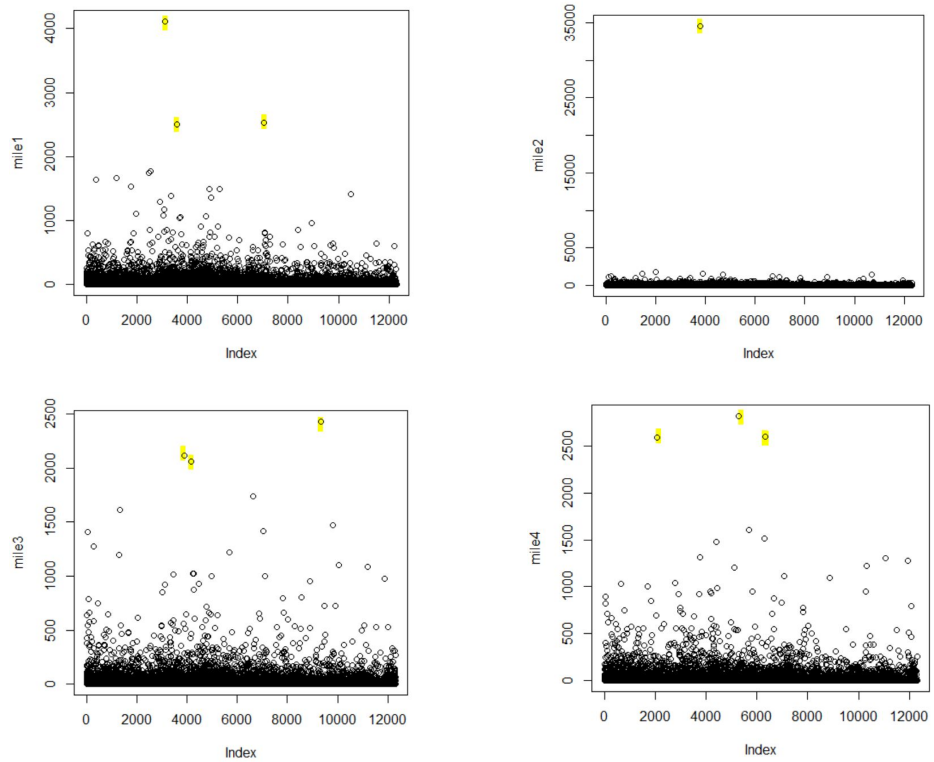
Correlations:

Prefood:basemile , Prebank:basemile and Elaboration:Particpate showed strong correlations indicating multicollinearity between these variables. Furthermore, we did a variance inflation factor test (in our regression models) that verified our assumption.

Correlation Matrix		
1	Prefood ~ basemile	0.69
2	Prebank ~ basemile	0.71
3	Elaboration ~ Particpate	0.77

Outliers:

Looking at the summary statistics and scatterplots of the independent variables we found certain extreme (in mile2) and few mild outliers (in mile1, mile3 and mile4) in the dataset:



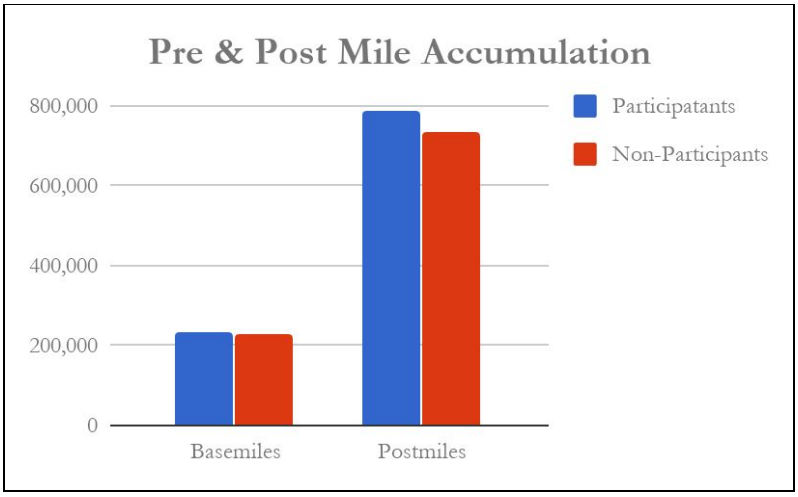
As natural as it is (and evident from mile1, mile3 and mile4 plots) some people got the most out of the campaign and accumulated a lot of miles. Mile2 outlier however, seems like typo error hence was removed from analysis.

Alteration to the database:

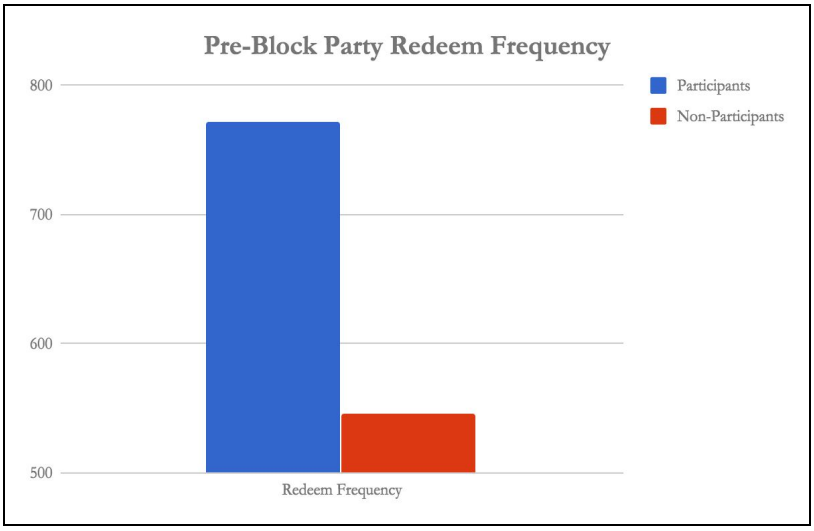
To measure cumulative effect on mile accumulation the after the party we created a column 'Postmiles' = Sum of mile1, mile2, mile3 and mile 4

Analysis:

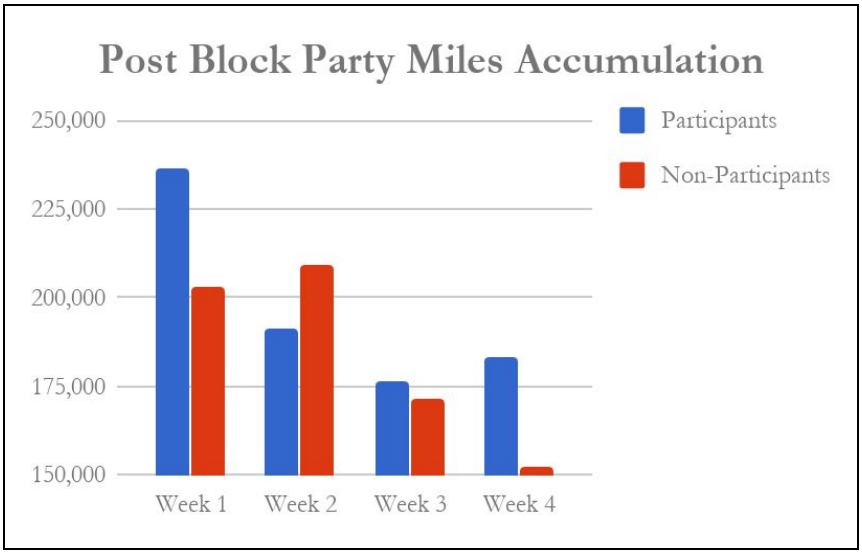
Below is comparison of the pre-Block Party mile accumulation vs. post-Block Party mile accumulation. As you can see, both participants (particpate = 1) and non-participants (particpate = 0) collected a similar total number of miles pre-party. A similar trend (however cumulatively larger) was observed in post-party miles between participants and non-participants.



Following is another graphic providing an overview of the 'good customers'. It shows that members who participated in the Block Party contest were typically higher redeemers of miles
- 772 total miles compared to 546 for non-participants.



Finally, the graph below displays the effect of the Block Party on the subsequent build up of miles. As you can see, the first week after the contest, there is a spike in mile accumulation with both participants and non-participants. Participants seem to be buying more miles than non-participants except for the 2nd week where non-participants surpass the participants by a margin of 209,291 to 191,112. In week 4, we see miles have trailed off from a high of 236,634 miles to 183,297 for participants; and from 209,291 to 151,976 for non-participants.



Does participation in the Block Party contest increase subsequent purchases?

Regression model:

We set out to analyse the variables responsible for the increase in mile accumulation after the Block Party contest. As mentioned earlier we saw strong correlations between Prefood:basemile , Prebank:basemile and Elaboration:Particpate. The largest collinearity observed was between "participate" and "elaboration". The correlation was ***r = .77***.

We regressed "mile1" on "participate", "elaboration", "RFreq", and "basemile". Output is show below

```
> round(cor(party),2)
>fit1=lm(log(mile1+1)~log(particpate+1)+log(elaboration+1)+log(Rfreq+1)+log(bas
emile), party)
> summary(fit1)
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           0.542006   0.030214  17.939   <2e-16 ***
log(particpate + 1)  -0.059993   0.121048  -0.496    0.620
log(elaboration + 1)  0.038778   0.027014   1.435    0.151
log(Rfreq + 1)        0.009411   0.059188   0.159    0.874
log(basemile)         0.509083   0.00axscd vfghuio=[-;p0o9i8u7yrfedx cnmkl.;p'[
}?.,m nbvcx8561  59.468   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.548 on 12287 degrees of freedom
Multiple R-squared:  0.2252,    Adjusted R-squared:  0.2249
F-statistic: 892.7 on 4 and 12287 DF,  p-value: < 2.2e-16
```

Hypothesis Test:

H0 : $\beta 1 = \beta 2 = \beta 3 = \beta 4 = 0$

H1 : at least one $\beta j \neq 0$

P-value < 2.2e-16, so the overall model is significant. But only basemile coefficient is significant in this model, the other three are not.

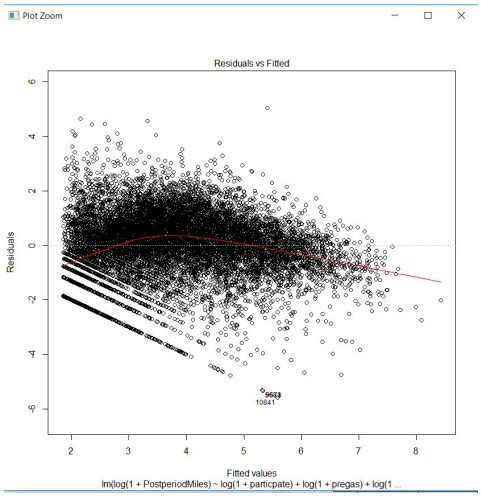
```
> vif(fit1)
 log(particpate + 1) log(elaboration + 1)       log(Rfreq + 1)        log(basemile)
         9.029604             9.034575             1.002196             1.004151
```

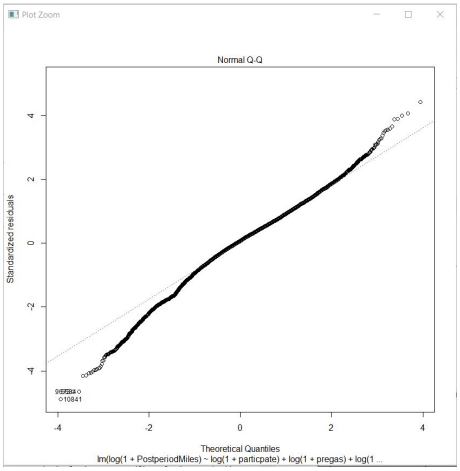
As vif is quite big for participate and elaboration (with 10 indicating a rather big multicollinearity problem). We decided to remove elaboration from the model. Updated model is as follows:

```
> fit1=lm(log(mile1+1)~log(particpate+1)+log(Rfreq+1)+log(basemile),party)
> summary(fit1)

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               0.54033    0.03019  17.896  < 2e-16 ***
log(particpate + 1)       0.10386    0.04029   2.578  0.00995 **
log(Rfreq + 1)            0.01061    0.05919   0.179  0.85769
log(basemile)             0.50970    0.00855  59.612  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.548 on 12288 degrees of freedom
Multiple R-squared:  0.225,     Adjusted R-squared:  0.2248
F-statistic:  1189 on 3 and 12288 DF,  p-value: < 2.2e-16
```

P-value < 2.2e-16, so the overall model is significant. Individual coefficients - particpate and basemile are also significant in this model, Rfreq is not significant though.

```
> plot(fit1)
```





<u>The model equation is:</u>

predicted log mile1 = 0.54 + 0.11log (participate+1) + 0.001log(Rfreq+1) + 0.51log (basemile)

Does the amount of elaboration affect subsequent purchases?

Regression model:

We have already unearthed the collinearity between the "particpate" and "elaboration" variables. Therefore, we do not need to conduct that analysis a second time. Based on the "Post Block Party Contest Miles Accumulation" histogram above, we decided to use "Week1" as the post-party variable/predictor in our regression. "Week1" experienced the highest miles accumulation.

```
> fit2=lm(log(mile1+1)~log(elaboration+1)+log(Rfreq+1)+log(basemile),party)
> summary(fit2)
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.539177   0.029668  18.174  < 2e-16 ***
log(elaboration + 1) 0.026153   0.008990   2.909  0.00363 **
log(Rfreq + 1)      0.009621   0.059185   0.163  0.87087
log(basemile)       0.509278   0.008551  59.556  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.548 on 12288 degrees of freedom
Multiple R-squared:  0.2252,    Adjusted R-squared:  0.225
F-statistic:  1190 on 3 and 12288 DF,  p-value: < 2.2e-16
```

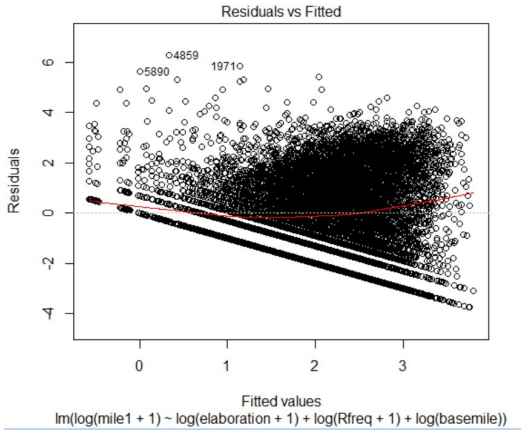
Hypothesis Test:

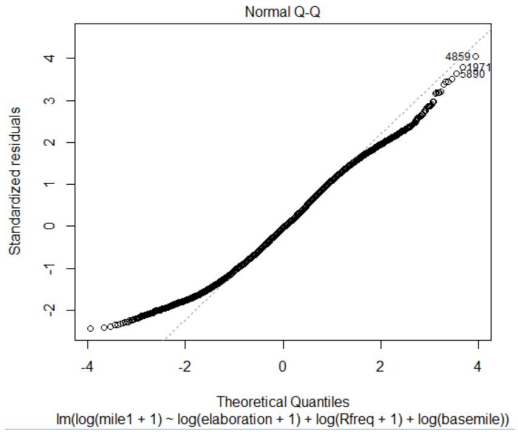
H0 : $\beta 1 = \beta 2 = \beta 3 = \beta 4 = 0$

H1 : at least one $\beta j \neq 0$

P-value < 2.2e-16, so the overall model is significant. Individual coefficients - elaboration and basemile are also significant in this model, Rfreq is not significant though.

```
> plot(fit2)
```





The model equation is:

predicted log mile1 = 0.54 + 0.03log (elaboration+1) + 0.01log(Rfreq+1) + 0.51log (basemile)

For how long does the participation/elaboration affect persist?

Regression model:

We built a linear model regressing mile1 - mile 4 (week1-week4) over pre measures as control variables for elaboration and participation. For participation, we concluded that significance of participation variable dropped over first 3 weeks and eventually saw to no significance by the end of fourth week. For elaboration, we had similar results over the first three weeks however, effect of elaboration persisted even in the fourth week.

Following are our findings from week 4.

For participation

```
> fit1=lm(log(1+mile4) ~ log(1+particpate)
+log(1+pregas)+log(1+prefood)+log(1+preother)+log(1+preretail)+log(1+prebank),party)
> summary(fit1)

Call:
lm(formula = log(1 + mile4) ~ log(1 + particpate) + log(1 + pregas) +
    log(1 + prefood) + log(1 + preother) + log(1 + preretail) +
    log(1 + prebank), data = party)

Residuals:
    Min      1Q  Median      3Q     Max
-4.2235 -1.0010 -0.2644  0.8625  6.6674

Coefficients:
                                    Estimate     Std. Error    t value
      Pr(>|t|)
      (Intercept)                   0.425324     0.027704      15.352       <
      2e-16 ***
      log(1 + particpate)        0.060431     0.037730     1.602        0.109
      log(1 + pregas)            0.224936     0.016874     13.331       < 2e-16 ***
      log(1 + prefood)           0.398482     0.009275     42.964       < 2e-16 ***
      log(1 + preother)          0.065265     0.016511     3.953        7.77e-05 ***
      log(1 + preretail)         0.127849     0.022009     5.809        6.45e-09 ***
      log(1 + prebank)           0.222404     0.008221     27.055       < 2e-16 ***
      ---
      Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

      Residual standard error: 1.449 on 12288 degrees of freedom
      Multiple R-squared:  0.2279,     Adjusted R-squared:  0.2275
      F-statistic: 604.8 on 6 and 12295 DF,  p-value: < 2.2e-16
```

For Elaboration:

```
> fit1=lm(log(1+mile4) ~ log(1+elaboration)
+log(1+pregas)+log(1+prefood)+log(1+preother)+log(1+preretail)+log(1+prebank),party)
       > summary(fit1)

Call:
lm(formula = log(1 + mile4) ~ log(1 + elaboration) + log(1 + pregas) + log(1 +
prefood) + log(1 + preother) + log(1 + preretail) + log(1 + prebank), data = party)

Residuals:
Min      1Q  Median      3Q     Max
-4.2301 -0.9994 -0.2629  0.8659  6.6736

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
       (Intercept)          0.419648   0.027210  15.422  < 2e-16 ***
       log(1 + elaboration) 0.018686   0.008419   2.219   0.0265 *
       log(1 + pregas)      0.224937   0.016872  13.332  < 2e-16 ***
       log(1 + prefood)     0.398261   0.009275  42.941  < 2e-16 ***
       log(1 + preother)    0.065337   0.016509   3.958 7.61e-05 ***
       log(1 + preretail)   0.127042   0.022011   5.772 8.04e-09 ***
       log(1 + prebank)     0.222294   0.008219  27.048  < 2e-16 ***
       ---
       Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

       Residual standard error: 1.449 on 12288 degrees of freedom
       Multiple R-squared:  0.228,      Adjusted R-squared:  0.2277
       F-statistic: 605.3 on 6 and 12295 DF,  p-value: < 2.2e-16
```

Conclusion and Summary

As evidenced by the data and proven through our analysis, it seems all members (especially, non-participants) were influenced by the Block Party contest.

It is recommended the Loyalty Program consider sponsoring such promotions in the future. The frequency and channel or format of future events is outside the scope of this report. However, it advisable that LP consider the continuation of a social media presence to allow member to communicate their purchase experiences.

Appendix

Summary of Independent variables:

```
     prefood             pregas            prebank
Min.   :  0.00     Min.   :  0.000     Min.   :  0.00
1st Qu.:  0.45     1st Qu.:  0.000     1st Qu.:  0.00
Median :  3.73     Median :  0.110     Median :  2.71
Mean   : 15.50      Mean   :  1.864    Mean   : 17.13
3rd Qu.: 14.65     3rd Qu.:  1.470     3rd Qu.: 22.13
Max.   :434.56     Max.   :245.560     Max.   :485.71

  preretail             preother                Rfreq
Min.   :  0.0000          Min.   :  0.0          Min.   : 0.0000
1st Qu.:  0.0000          1st Qu.:  0.0           1st Qu.: 0.0000
Median :  0.0000          Median :  0.0          Median : 0.0000
Mean   :  0.9898    Mean   :  2.1          Mean   : 0.1071
3rd Qu.:  0.3400          3rd Qu.:  0.9          3rd Qu.: 0.0000
Max.   :403.1800    Max.   :371.7          Max.   :63.0000

    mile1              mile2            mile3
Min.   :   0.00    Min.   :    0.00    Min.   :   0.00
1st Qu.:   0.00    1st Qu.:    0.00    1st Qu.:   0.00
Median :   4.00    Median :    4.00    Median :   4.00
Mean   :  35.73    Mean   :   32.55    Mean   :  28.32
3rd Qu.:  25.00    3rd Qu.:   19.00    3rd Qu.:  20.00
Max.   :4117.00    Max.   :34576.00    Max.   :2435.00

    mile4            basemile            postmile
Min.   :   0.00    Min.   :  0.11      Min.   :    0.0
1st Qu.:   0.00    1st Qu.:  5.65      1st Qu.:   13.0
Median :   3.00    Median : 18.63      Median :   50.0
Mean   :  27.25    Mean   : 37.58      Mean   :  123.8
3rd Qu.:  14.00    3rd Qu.: 49.34      3rd Qu.:  148.0
Max.   :2821.00    Max.   :498.81      Max.   :34771.0

 elaboration
Min.   : 0.00
1st Qu.: 0.00
Median : 0.00
Mean   :10.99
3rd Qu.:20.00
Max.   :56.00
```

Correlation matrix of Independent variable:

	X	particpate	elaboration	prefood	pregas
X	1.00000000	0.027312628	0.007768180	-0.11401803	-0.061557879
particpate	0.02731263	1.000000000	0.770942426	0.02769220	-0.003394842
elaboration	0.00776818	0.770942426	1.000000000	0.03625407	0.009332453
prefood	-0.11401803	0.027692197	0.036254075	1.00000000	0.039978585
pregas	-0.06155788	-0.003394842	0.009332453	0.03997859	1.000000000
prebank	-0.19846639	-0.017585797	0.002977255	0.06869188	0.155284267
preretail	-0.02611311	0.030145120	0.034453046	0.02979194	0.017371972
preother	-0.04751666	-0.015984535	-0.008693039	0.02990948	0.026019679
Rfreq	-0.02437340	0.014642463	0.019331504	0.04804456	-0.001740324
mile1	-0.11212680	0.025771690	0.036201008	0.30581770	0.054864637
mile2	-0.03280973	-0.004586424	-0.001304486	0.10339960	0.020609529
mile3	-0.07604723	0.004665831	0.005649404	0.23134975	0.083650844
mile4	-0.08649865	0.027871250	0.037756466	0.21582911	0.074403589
basemile	-0.21542966	0.006568136	0.028178678	0.69430950	0.229014584
Postmiles	-0.09904754	0.011290373	0.019742847	0.28277515	0.070985594

	prebank	preretail	preother	Rfreq	mile1
X	-0.198466391	-0.026113113	-0.047516662	-0.024373401	-0.11212680
particpate	-0.017585797	0.030145120	-0.015984535	0.014642463	0.02577169
elaboration	0.002977255	0.034453046	-0.008693039	0.019331504	0.03620101
prefood	0.068691883	0.029791944	0.029909481	0.048044560	0.30581770
pregas	0.155284267	0.017371972	0.026019679	-0.001740324	0.05486464
prebank	1.000000000	0.046985310	0.123030609	0.031083324	0.22914765
preretail	0.046985310	1.000000000	0.018436873	-0.003096389	0.02447277
preother	0.123030609	0.018436873	1.000000000	0.003114944	0.06651838
Rfreq	0.031083324	-0.003096389	0.003114944	1.000000000	0.01331966
mile1	0.229147648	0.024472769	0.066518378	0.013319663	1.00000000
mile2	0.067581886	0.007423585	0.014590932	0.007891530	0.02590958
mile3	0.302722625	0.054588643	0.036551605	0.024269406	0.07026008
mile4	0.265565131	0.015688911	0.065020854	0.020408866	0.06280424
basemile	0.711000335	0.177409942	0.324982541	0.049989407	0.36006861
Postmiles	0.258672506	0.029853405	0.055925340	0.021215702	0.33924094

	mile2	mile3	mile4	basemile	Postmiles
X	-0.032809730	-0.076047232	-0.08649865	-0.215429663	-0.09904754
particpate	-0.004586424	0.004665831	0.02787125	0.006568136	0.01129037
elaboration	-0.001304486	0.005649404	0.03775647	0.028178678	0.01974285
prefood	0.103399599	0.231349751	0.21582911	0.694309498	0.28277515
pregas	0.020609529	0.083650844	0.07440359	0.229014584	0.07098559
prebank	0.067581886	0.302722625	0.26556513	0.711000335	0.25867251
preretail	0.007423585	0.054588643	0.01568891	0.177409942	0.02985341
preother	0.014590932	0.036551605	0.06502085	0.324982541	0.05592534
Rfreq	0.007891530	0.024269406	0.02040887	0.049989407	0.02121570
mile1	0.025909580	0.070260085	0.06280424	0.360068608	0.33924094
mile2	1.000000000	0.032926251	0.02212657	0.113947127	0.88333501
mile3	0.032926251	1.000000000	0.06520199	0.357930435	0.29575061
mile4	0.022126571	0.065201987	1.00000000	0.325744164	0.29667552
basemile	0.113947127	0.357930435	0.32574416	1.000000000	0.36352021
Postmiles	0.883335014	0.295750610	0.29667552	0.363520205	1.00000000