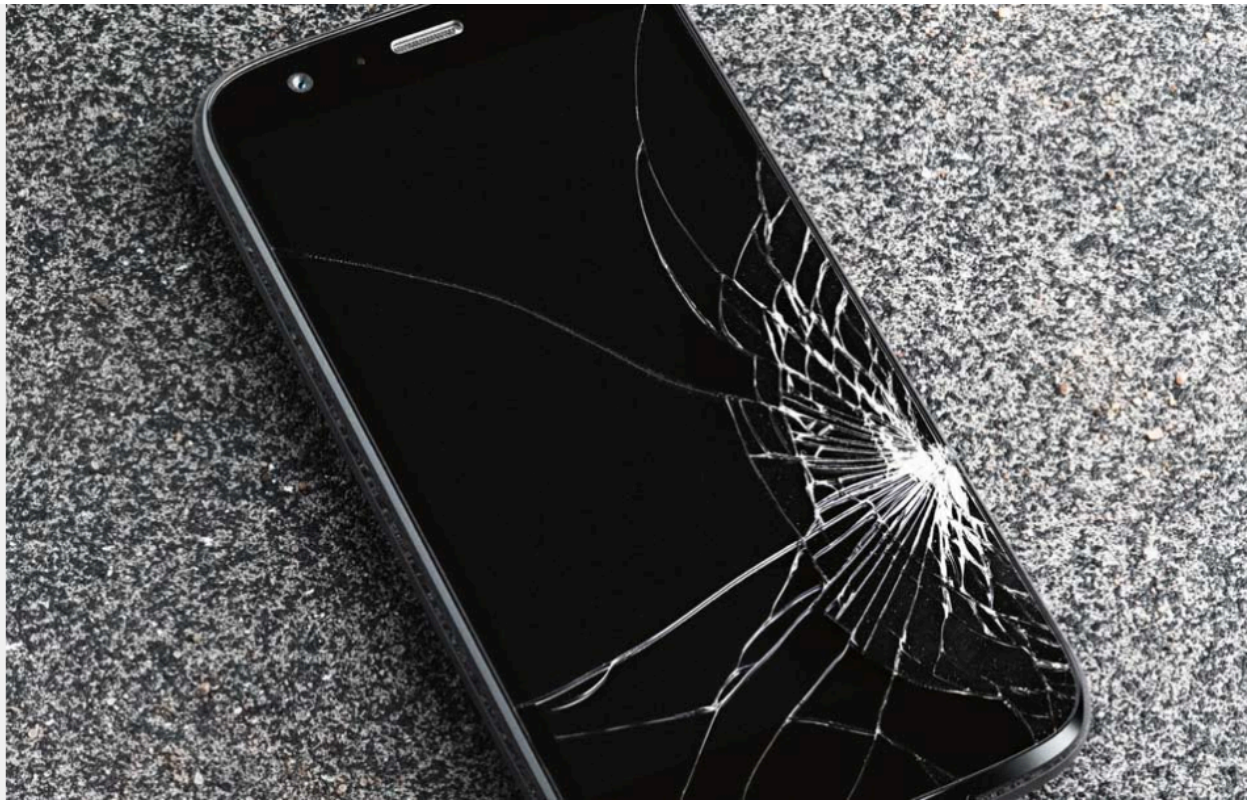
Case Study 3

Best Buy Warranty Sales Project

Presenters: Katharine Grant, Ren Hui, Xin Zou, Yezhu Li



Outline

- ✔ *Intro & Data Summary*
- ✔ *Model Verification*
- ✔ *Results Interpretation*
- ✔ *Proposal & Limitations*

Introduction

Background Recap

Warranty, is a Geek Squad protection plan purchased along with a product. The Geek Squad protection plan aims to insure customers after the manufacturer's warranty expires. Warranty sale becomes an important marketing strategy and offers substantial profit for Best Buy.

Research Purpose

To identify customers who have a high propensity to purchase warranty;

To make a data-driven proposal for the potential marketing strategy.

Data Snapshot

personid	*age*	*hisp*	*PriceCategory*	*married*	*MyBestBuy*	*hhincome*	*appliances*	*Warranty*	*familysize*	*productgeneration*	*newcustomer*	*weekend*
54963010	62	0	12	0	0	0	0	0	2	7	1	1
21629010	59	0	12	0	0	0	0	1	2	7	1	1
20421010	60	0	13	0	1	0	1	0	1	8	0	0
38784010	62	0	10	0	0	0	0	1	2	6	1	1
55630030	54	0	9	0	0	0	0	1	2	5	0	0
15893020	62	0	12	1	1	0	1	0	4	7	0	0
86263010	59	0	5	1	0	0	0	1	3	4	1	0
51270010	59	0	11	0	0	0	0	0	2	7	1	0

***3,206 transactions, 13 Variables**, made in March 2017 from Santa Clara Best Buy stores.*

Variables of Theoretical Model

Dependent Variables

- ***Warranty*** *= 1 if customer i purchased a Geek Squad protection plan along with the product and 0 otherwise*

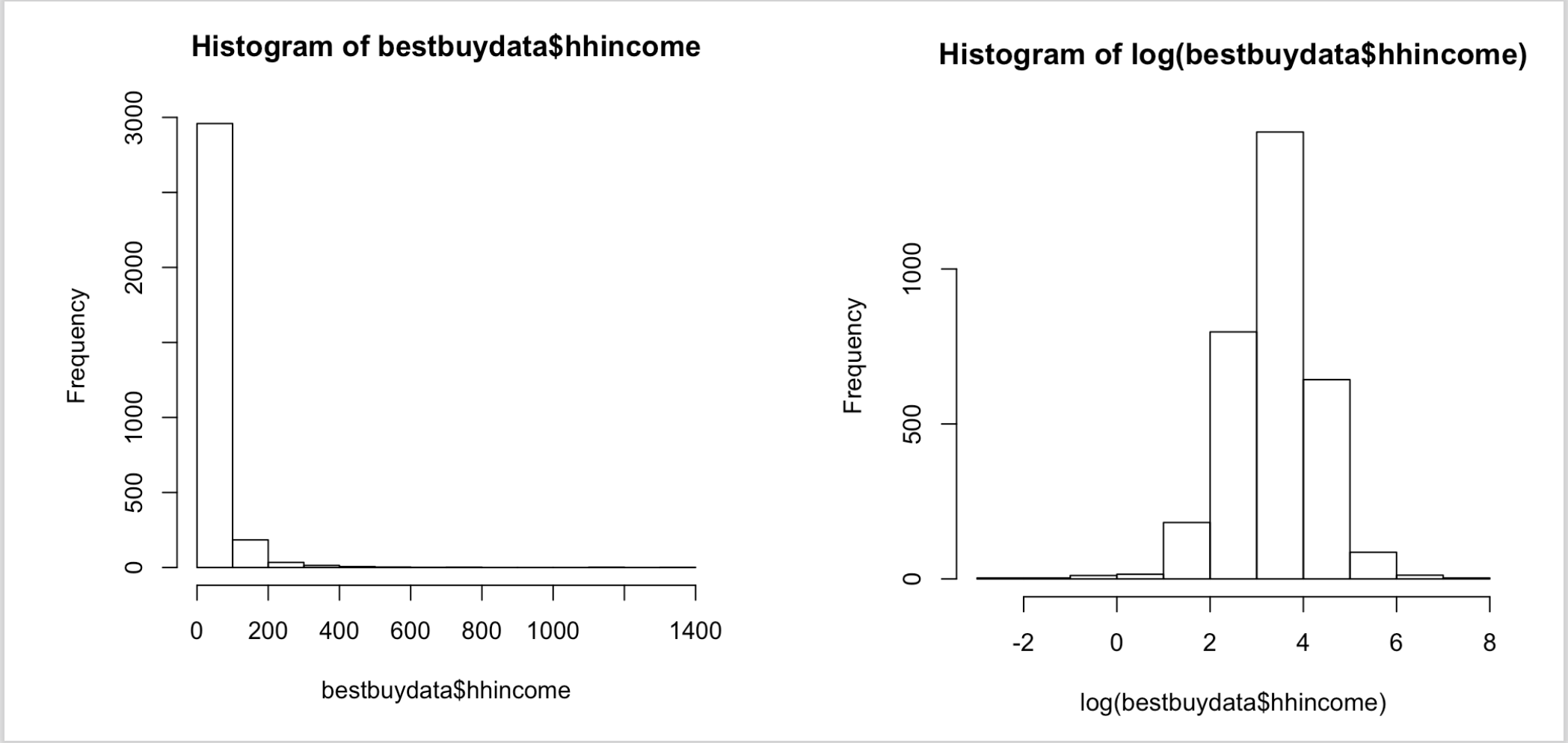
Variables of Interest

- ***PriceCategory*** *= categorical variable between 0 and 17 that defines the value of the purchased product.*

Other Variables of Interest

- ***MyBestBuy*** *= 1 if customer i has MyBestBuy (i.e. company) credit card and 0 otherwise*
- ***Appliances*** *= 1 if customer i purchased a product from the "home appliances" category and 0 otherwise*
- ***ProductGeneration*** *= indicates the generation of a product within its product domain. A high number indicates a new generation.*
- ***Demographic: Age, Hisp, Married, Hhincome, Familysize***

Log-transformation



The Theoretical Model

- ✔ *Logit*

$$
\begin{aligned}
&Warranty_i \\
&= \beta_0 + \beta_1 PriceCategory_i + \beta_2 Appliances_i + \beta_3 log(1 + hhincome)_i \\
&+ \beta_4 ProducteGeneration_i + \beta_5 MyBestBuy_i + \beta_6 Hisp_i + \beta_7 Age_i \\
&+ \beta_8 Married_i + \beta_9 Familysize_i + \varepsilon_i
\end{aligned}
$$

Model Verification

Multicollinearity

```
> vif(df1)
                          Variables      VIF
1             bestbuydata.Warranty 1.093986
2         bestbuydata.PriceCategory 9.622327
3   bestbuydata.productgeneration 9.228072
4             bestbuydata.MyBestBuy 1.136930
5           bestbuydata.appliances 1.192873
6                  bestbuydata.age 1.169291
7              bestbuydata.married 4.577358
8             bestbuydata.hhincome1 1.670485
9                  bestbuydata.hisp 1.160839
10          bestbuydata.familysize 4.221931
```

```
> vif(df3)
                          Variables      VIF
1             bestbuydata.Warranty 1.086805
2         bestbuydata.PriceCategory 1.444348
3             bestbuydata.MyBestBuy 1.135174
4           bestbuydata.appliances 1.192517
5                  bestbuydata.age 1.166522
6           bestbuydata.familysize 1.231464
7             bestbuydata.hhincome1 1.564901
8                  bestbuydata.hisp 1.160560
```

Interaction term

```
Call:
glm(formula = Warranty ~ age + hisp + PriceCategory + MyBestBuy +
    hhincome1 + appliances + familysize + PriceCategory:appliances,
    family = binomial, data = bestbuydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0344  -1.1836   0.7358   0.9121   2.1384

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               -1.45039    0.78212  -1.854   0.0637 .
age                       -0.01844    0.01139  -1.619   0.1055
hisp                      -1.39394    0.16442  -8.478  < 2e-16 ***
PriceCategory              0.13194    0.02327   5.670 1.43e-08 ***
MyBestBuy                  0.15049    0.08420   1.787   0.0739 .
hhincome1                  0.25545    0.05432   4.703 2.57e-06 ***
appliances                 2.74865    0.31664   8.681  < 2e-16 ***
familysize                 0.34960    0.04062   8.606  < 2e-16 ***
PriceCategory:appliances -0.24913    0.02758  -9.033  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4255.7  on 3205  degrees of freedom
Residual deviance: 3909.7  on 3197  degrees of freedom
AIC: 3927.7

Number of Fisher Scoring iterations: 4
```

***Theory**:*

- *More expensive appliances come with longer warranties*
- *Less likely that consumers will buy Best Buy warranty*

Interaction term tests

```
> AIC(logit2,logit3)
       df      AIC
logit2  8 4014.789
logit3  9 3927.695
> BIC(logit2,logit3)
       df      BIC
logit2  8 4063.371
logit3  9 3982.350
```

```
> with(logit2, pchisq(null.deviance - deviance, df.null
- df.residual, lower.tail = FALSE))
[1] 9.177856e-52
> with(logit3, pchisq(null.deviance - deviance, df.null
- df.residual, lower.tail = FALSE))
[1] 6.322924e-70
```

```
> anova(logit2, logit3, test="Chisq")
Analysis of Deviance Table

Model 1: Warranty ~ age + hisp + PriceCategory + MyBestBuy + hhincome1 +
    appliances + familysize
Model 2: Warranty ~ age + hisp + PriceCategory + MyBestBuy + hhincome1 +
    appliances + familysize + PriceCategory:appliances
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      3198     3998.8
2      3197     3909.7  1   89.094 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Heteroskedasticity

- *There is heteroskedasticity*
- *The coefficients and p-values do not change when we include robust standard errors*

```
> gqtest(logit3) # p-value is insignificant

        Goldfeld-Quandt test

data:  logit3
GQ = 0.97535, df1 = 1594, df2 = 1594, p-value = 0.6908
alternative hypothesis: variance increases from segment 1 to 2

> bptest(logit3) # p-value is significant

        studentized Breusch-Pagan test

data:  logit3
BP = 53.831, df = 8, p-value = 7.445e-09
```

```
> coeftest(logit3, vcov = vcovHC(logit3, "HC1"))

z test of coefficients:

                           Estimate Std. Error z value  Pr(>|z|)
(Intercept)               -1.450389   0.788276 -1.8400   0.06578 .
age                       -0.018438   0.011505 -1.6026   0.10901
hisp                      -1.393938   0.160750 -8.6715 < 2.2e-16 ***
PriceCategory              0.131937   0.022821  5.7814 7.408e-09 ***
MyBestBuy                  0.150488   0.084918  1.7722   0.07637 .
hhincome1                  0.255453   0.054002  4.7304 2.240e-06 ***
appliances                 2.748647   0.312145  8.8057 < 2.2e-16 ***
familysize                 0.349603   0.040773  8.5744 < 2.2e-16 ***
PriceCategory:appliances -0.249134   0.027057 -9.2078 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Final Model

Logit

$$Warranty_i = -1.45 + 0.13PriceCategory_i + 2.75Appliances_i + 0.26log(1 + hhincome)_i + 0.15MyBestBuy_i - 1.39Hisp_i - 0.02age_i + 0.35Familysize_i - 0.25PriceCategory_i * Appliances_i$$

Probit

$$Warranty_i = -0.89 + 0.08PriceCategory_i + 1.66Appliances_i + 0.16log(1 + hhincome)_i + 0.09MyBestBuy_i - 0.85Hisp_i - 0.01age_i + 0.21Familysize_i - 0.15PriceCategoryi * Appliances_i$$

Probit Model

```
Call:
glm(formula = Warranty ~ age + hisp + PriceCategory + MyBestBuy +
    hhincome1 + appliances + familysize + PriceCategory:appliances,
    family = binomial(link = "probit"), data = bestbuydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0591  -1.1836   0.7373   0.9160   2.1812

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              -0.888531   0.471974  -1.883   0.0598 .
age                      -0.011061   0.006886  -1.606   0.1082
hisp                     -0.851901   0.098374  -8.660  < 2e-16 ***
PriceCategory             0.079961   0.013718   5.829 5.58e-09 ***
MyBestBuy                 0.092168   0.051111   1.803   0.0713 .
hhincome1                 0.155066   0.032879   4.716 2.40e-06 ***
appliances                1.658999   0.188129   8.818  < 2e-16 ***
familysize                0.213400   0.024731   8.629  < 2e-16 ***
PriceCategory:appliances -0.150419   0.016286  -9.236  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4255.7  on 3205  degrees of freedom
Residual deviance: 3909.7  on 3197  degrees of freedom
AIC: 3927.7

Number of Fisher Scoring iterations: 4
```

Final Logit Model

$$
\begin{aligned}
&Warranty_i \\
&= -1.45 + 0.13PriceCategory_i + 2.75Appliances_i \\
&+ 0.26log(1 + hhincome)_i + 0.15MyBestBuy_i - 1.39Hisp_i \\
&- 0.02age_i + 0.35Familysize_i + \varepsilon_i
\end{aligned}
$$

Final Model

```
Call:
glm(formula = Warranty ~ age + hisp + PriceCategory + MyBestBuy +
    hhincome1 + appliances + familysize + PriceCategory:appliances,
    family = binomial, data = bestbuydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0344  -1.1836   0.7358   0.9121   2.1384

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               -1.45039    0.78212  -1.854   0.0637 .
age                       -0.01844    0.01139  -1.619   0.1055
hisp                      -1.39394    0.16442  -8.478  < 2e-16 ***
PriceCategory              0.13194    0.02327   5.670 1.43e-08 ***
MyBestBuy                  0.15049    0.08420   1.787   0.0739 .
hhincome1                  0.25545    0.05432   4.703 2.57e-06 ***
appliances                 2.74865    0.31664   8.681  < 2e-16 ***
familysize                 0.34960    0.04062   8.606  < 2e-16 ***
PriceCategory:appliances -0.24913    0.02758  -9.033  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4255.7  on 3205  degrees of freedom
Residual deviance: 3909.7  on 3197  degrees of freedom
AIC: 3927.7

Number of Fisher Scoring iterations: 4
```

Interpretation of results

The likelihood ratio test results shows that our model has a good fit!

```
> with(logit7,null.deviance-deviance)
[1] 346.0453
> with(logit7,df.null-df.residual)
[1] 8
> with(logit7,pchisq(null.deviance-deviance,df.null-df.residual,lower.tail = FALSE))
[1] 6.322924e-70
```

Interpretation of results

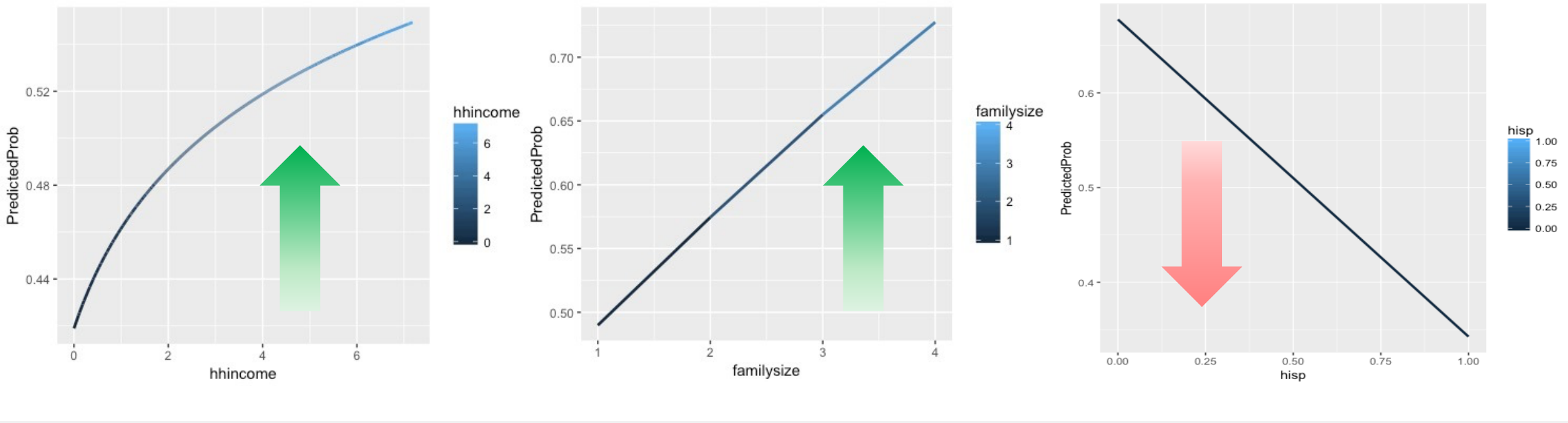
Our model has a 67% correct Classification rate!


```
> pred<-predict(logit7,data=BestBuy,type="response")
> prediction<-ifelse(pred>=0.5,1,0)
> error<-mean(prediction!=BestBuy$Warranty)
> 1-error
[1] 0.6746725
```

Interpretation of results(individual variable)

Variables	*Relationship with buying warranty*	β	e^{β}	*Marginal Effects*
Household Income	Positive	0.255(log)	1.290(log)	0.059%
Family size	Positive	0.350	1.419	8%
Hispanic	Negative	-1.394	0.248	-33%
Price Category	Interaction	See next slide		
Appliances	Interaction			
Age	unknown	NA		
MyBestBuy	unknown	NA		

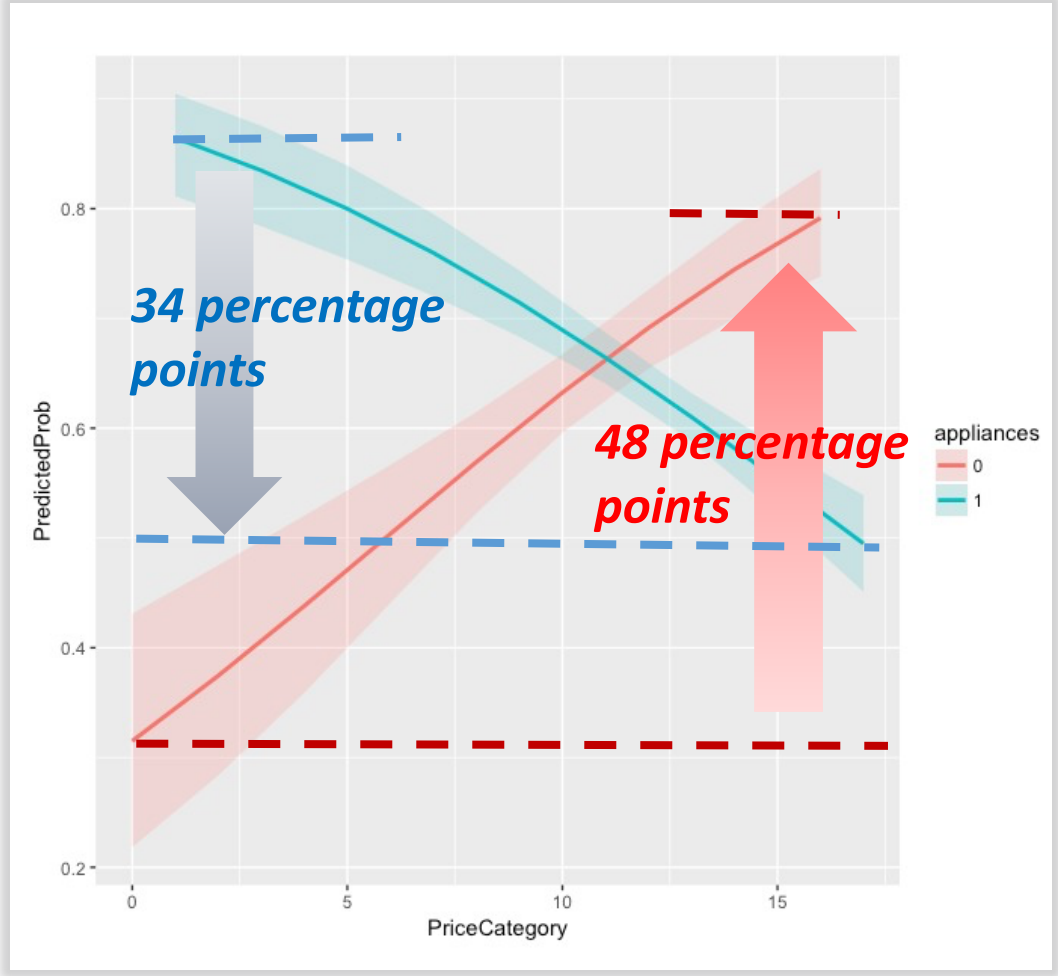
Interpretation of results(individual variable)



Interpretation of results(interaction term)

Holding other variables constant…

Price Category	*Non-Appliances*	*Appliances*
β	0.132	-0.117(=0.132-0.249)
e^{β}(odds ratio)	1.141	0.890
Marginal Effects	3%	-2%



Interpretation of results(individual variable)

Variables	*Relationship with buying warranty*	β	e^{β}	*Marginal Effects*
Household Income	Positive	0.255(log)	1.290(log)	0.059%
Family size	Positive	0.350	1.419	8%
Hispanic	Negative	-1.394	0.248	-33%
Price Category	Interaction	See next slide		
Appliances	Interaction			
Age	unknown	NA		
MyBestBuy	unknown	NA		

Implications for Marketing Strategy

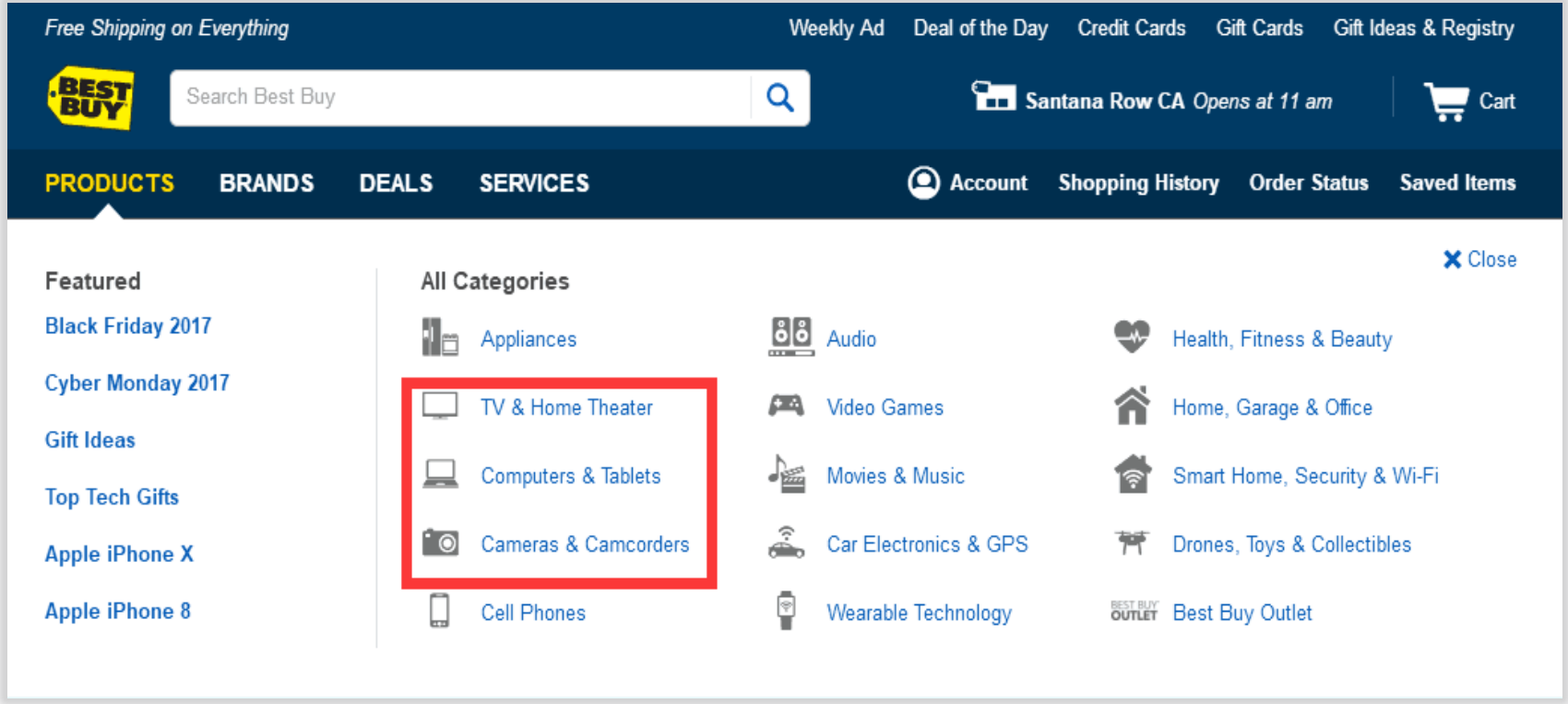
✔ ***The important things to decide our target***

- *No. 1: Non-appliances with higher price*
- *No. 2: Appliances with lower price*

- *No. 3: Non-Hispanic*
- *No. 4: Larger Family size*
- *No. 5: Higher income*

Implications for Marketing Strategy

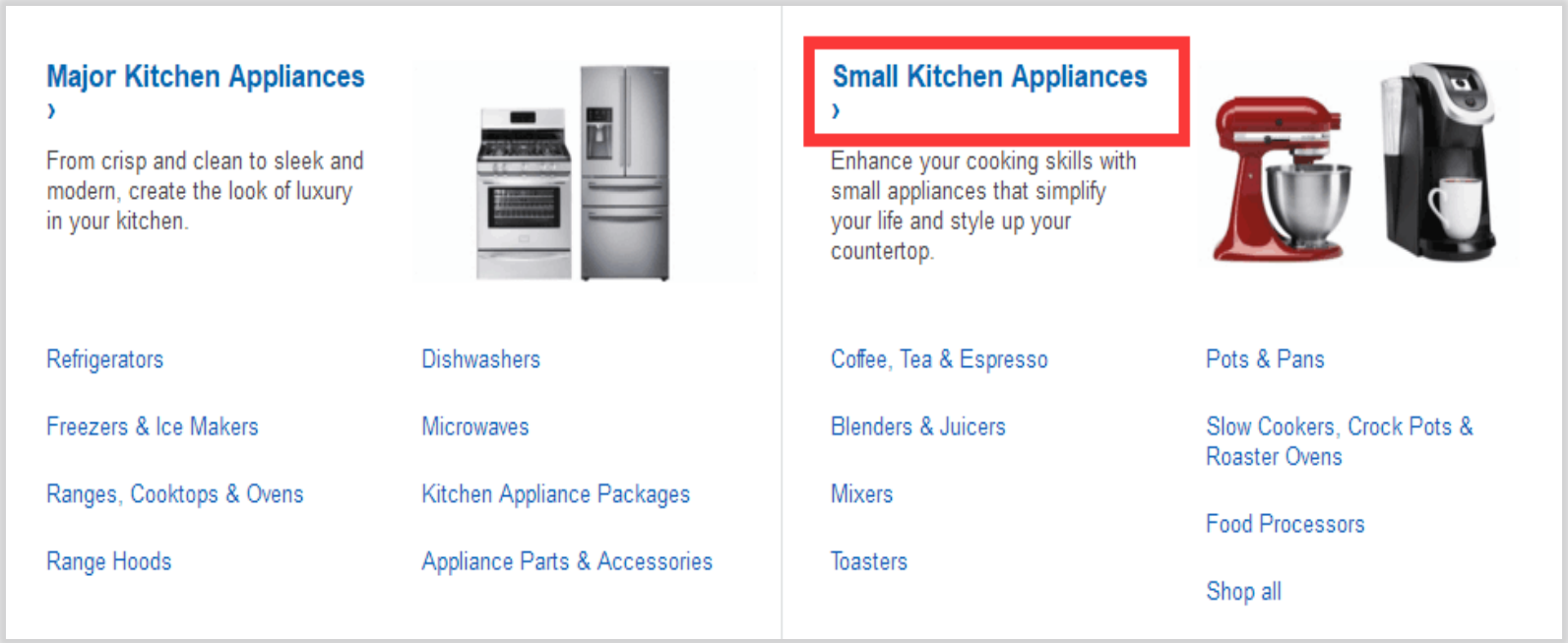
- *No. 1: Non-appliances with higher price*



Target Customers: Likely to decide company supplies, buy new house, want tech for recreation.

Implications for Marketing Strategy

- *No. 2: Appliances with lower price*



Target Customers: High tendency to make decisions for buying small appliances and home improvement, focused on convenience.

Implications for Marketing Strategy

- *No. 3: Non-Hispanic*

Target Customers: English speakers

- *No. 4: Larger Family size*

Target Customers: Family-oriented

- *No. 5: Higher income*

Target Customers: Business-oriented

Marketing Plan Proposal

- *Sending follow-up emails for TV/Home Theater, Computer, Camera, Small kitchen appliances buyers*

- *Increase Social Media marketing through influencers for family-oriented customers*

Online

- *Releasing ads at transportation such as airport for Business-oriented customers*

Offline

- *Won't increase budget for Spanish-language promotion, only increase English ads*

Limitation

1. *Some variables are not convincing.*
2. *Narrow consumer data scale.*
3. *Limitation of demographic info.*

Thanks!

Correlation

```
> cor(mydata)
                    personid           age          hisp PriceCategory     married
personid         1.000000000 -0.0330527947  0.0967204647  -0.048153525 -0.03817971
age             -0.033052795  1.0000000000 -0.0006346689   0.039916952  0.15756214
hisp             0.096720465 -0.0006346689  1.0000000000  -0.328141800 -0.02658656
PriceCategory   -0.048153525  0.0399169525 -0.3281418001   1.000000000  0.07494838
married         -0.038179713  0.1575621439 -0.0265865550   0.074948382  1.00000000
MyBestBuy       -0.032355211  0.3115567007 -0.0659347592   0.135125725  0.12202874
hhincome        -0.018325405  0.0652357371 -0.1052860258   0.290427681  0.20854885
appliances      -0.032671960  0.1967221171 -0.1163422707   0.309767740  0.09605320
Warranty        -0.021868583  0.0129989178 -0.1847920169   0.071934974  0.22135089
familysize      -0.036133232  0.1456522428 -0.0147721428   0.064004686  0.87308567
productgeneration -0.027111898  0.0542402527 -0.3080954399   0.944045453  0.07437216
newcustomer     -0.002354919  0.0120054043  0.0134669523  -0.005186286  0.01522605
weekend         -0.014707124 -0.0083913126 -0.0149751267  -0.011490043  0.03270994
                   MyBestBuy    hhincome   appliances    Warranty  familysize
personid        -0.032355211 -0.01832540 -0.032671960 -0.02186858 -0.03613323
age              0.311556701  0.06523574  0.196722117  0.01299892  0.14565224
hisp            -0.065934759 -0.10528603 -0.116342271 -0.18479202 -0.01477214
PriceCategory    0.135125725  0.29042768  0.309767740  0.07193497  0.06400469
married          0.122028744  0.20854885  0.096053202  0.22135089  0.87308567
MyBestBuy        1.000000000 -0.01204500  0.128009299  0.05936840  0.10869629
hhincome        -0.012045000  1.00000000  0.155125760  0.10770128  0.17158996
appliances       0.128009299  0.15512576  1.000000000  0.04482788  0.06402652
Warranty         0.059368397  0.10770128  0.044827879  1.00000000  0.19979141
familysize       0.108696288  0.17158996  0.064026525  0.19979141  1.00000000
productgeneration  0.133728765  0.28408757  0.293973195  0.06093647  0.06164029
newcustomer     -0.047584715  0.01221961 -0.002667570  0.01075128  0.02360174
weekend          0.005156923  0.01489175  0.007693935  0.02250629  0.01707987
```

Correlation

	productgeneration	newcustomer	weekend
personid	-0.02711190	-0.002354919	-0.014707124
age	0.05424025	0.012005404	-0.008391313
hisp	-0.30809544	0.013466952	-0.014975127
PriceCategory	0.94404545	-0.005186286	-0.011490043
married	0.07437216	0.015226049	0.032709938
MyBestBuy	0.13372877	-0.047584715	0.005156923
hhincome	0.28408757	0.012219606	0.014891748
appliances	0.29397319	-0.002667570	0.007693935
Warranty	0.06093647	0.010751281	0.022506286
familysize	0.06164029	0.023601740	0.017079870
productgeneration	1.00000000	-0.011593103	-0.014526542
newcustomer	-0.01159310	1.000000000	0.010458271
weekend	-0.01452654	0.010458271	1.000000000

Probit model predict accuracy rate

```
> pred<-predict(probit1,data=BestBuy,type="response")
> predic<-ifelse(pred>=0.5,1,0)
> error<-mean(predic!=BestBuy$Warranty)
> 1-error
[1] 0.6752963
```