

Customer Purchasing Behavior Analysis

Usman Khan

Metro College



Agenda

1 About Dataset

2 Business Questions

3 Exploratory Data Analysis

4 Insights and Trends

5 Takeways



The data I have...

- Customer Purchasing Behavior
- Observations: 1966
- Variables: 49
- Location: US





Research Question

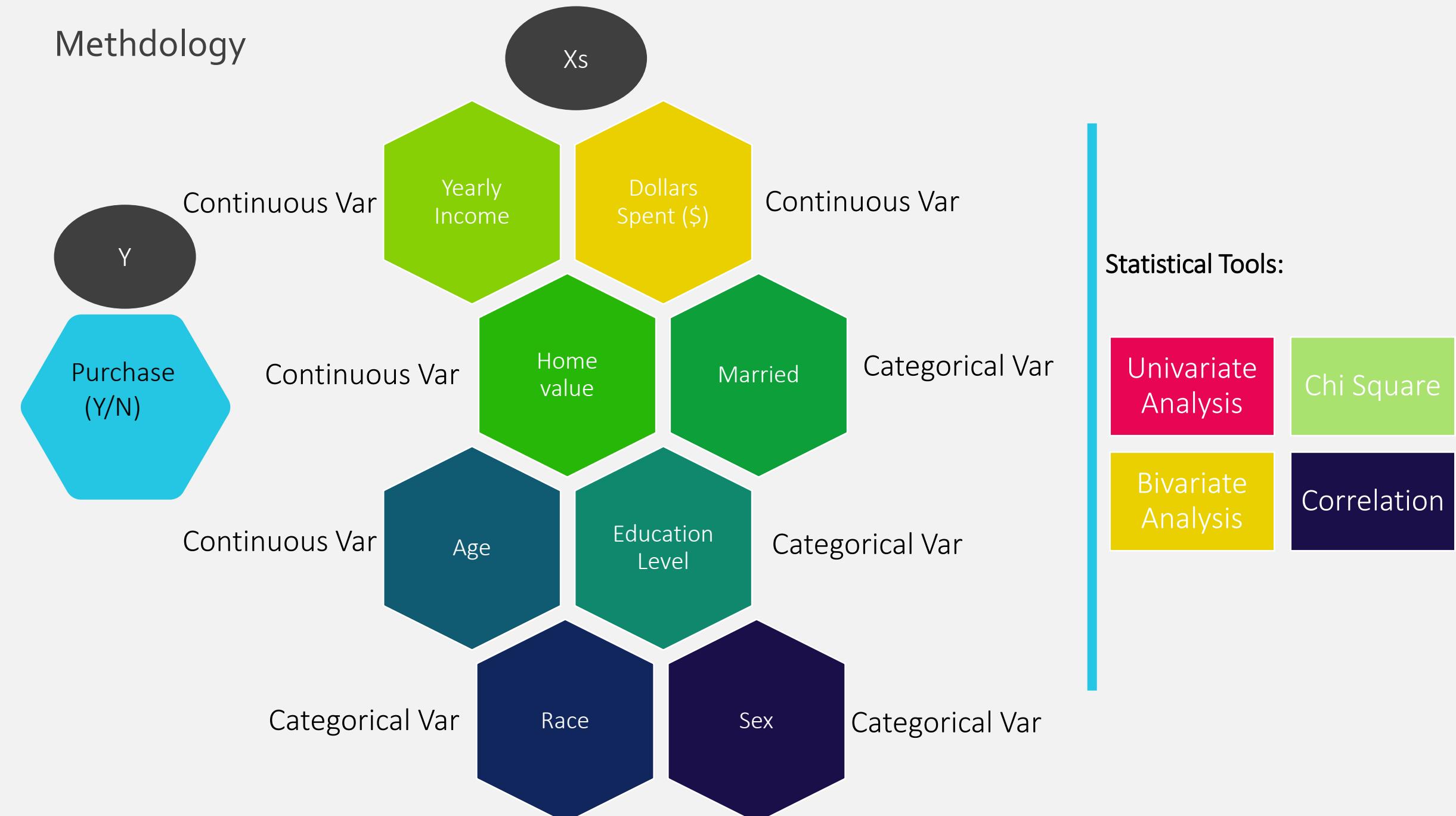
What are the trends in the data?



Business Questions

- Who buys our product the most?
- Who buys less and why?
- What are the takeaways?

Methodology



Frequency
Percent
Row Pct
Col Pct

		Table of PURCHASE by MARITAL		
		MARITAL(Married (y/n))		
PURCHASE(Purchase (y/n))		0	1	Total
0	0	291	676	967
	1	14.80	34.38	49.19
		30.09	69.91	
		45.47	50.98	
1	0	349	650	999
	1	17.75	33.06	50.81
		34.93	65.07	
		54.53	49.02	
Total		640	1326	1966
		32.55	67.45	100.00

Statistics for Table of PURCHASE by MARITAL

Statistic	DF	Value	Prob
Chi-Square	1	5.2466	0.0220
Likelihood Ratio Chi-Square	1	5.2524	0.0219
Continuity Adj. Chi-Square	1	5.0284	0.0249
Mantel-Haenszel Chi-Square	1	5.2439	0.0220
Phi Coefficient		-0.0517	
Contingency Coefficient		0.0516	
Cramer's V		-0.0517	

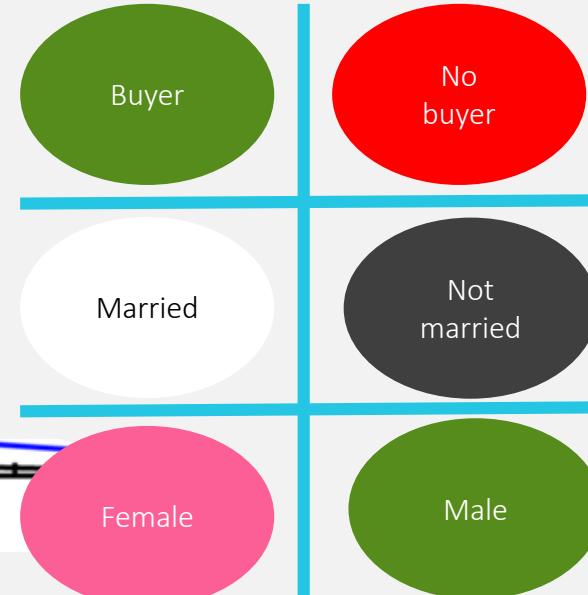
EDA - Chi-Square

Purchase vs Marital Status

Fisher's Exact Test

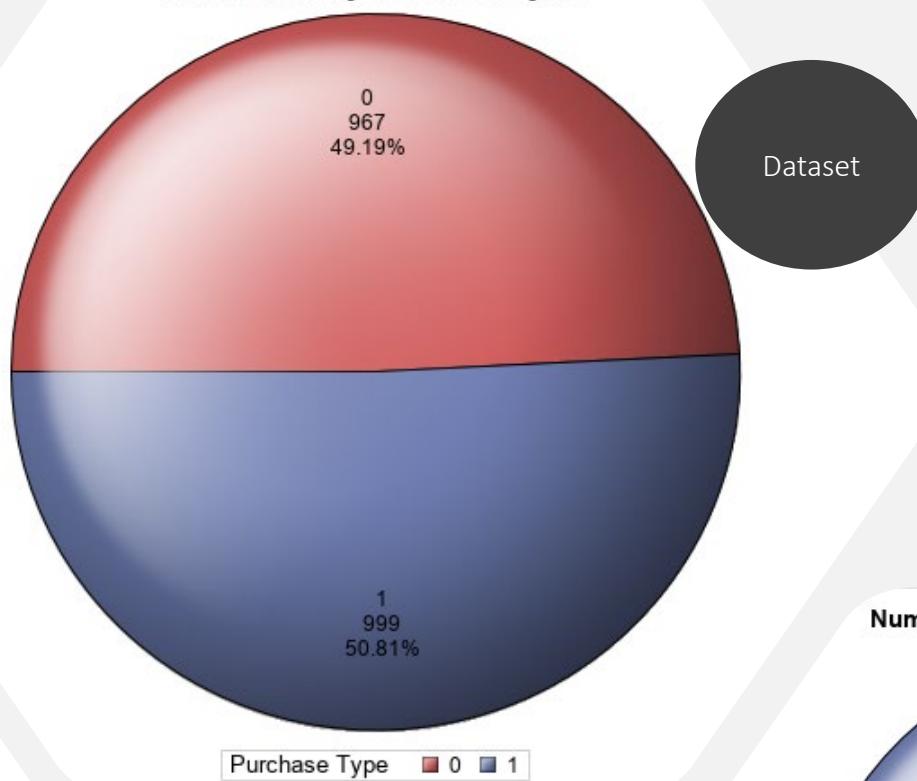
Cell (1,1) Frequency (F)	291
Left-sided Pr <= F	0.0124
Right-sided Pr >= F	0.9904
Table Probability (P)	0.0028
Two-sided Pr <= P	0.0237

Important Categories



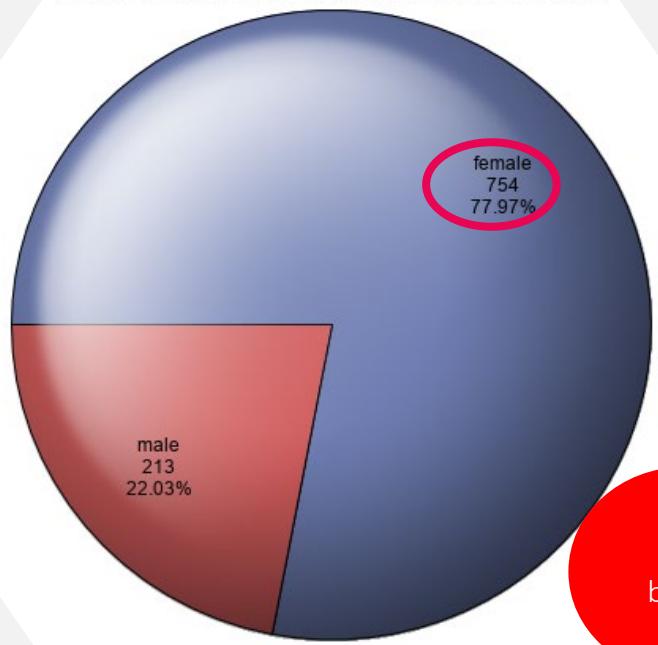
Who is buying? (Gender)

Number of Buyers vs Nonbuyers



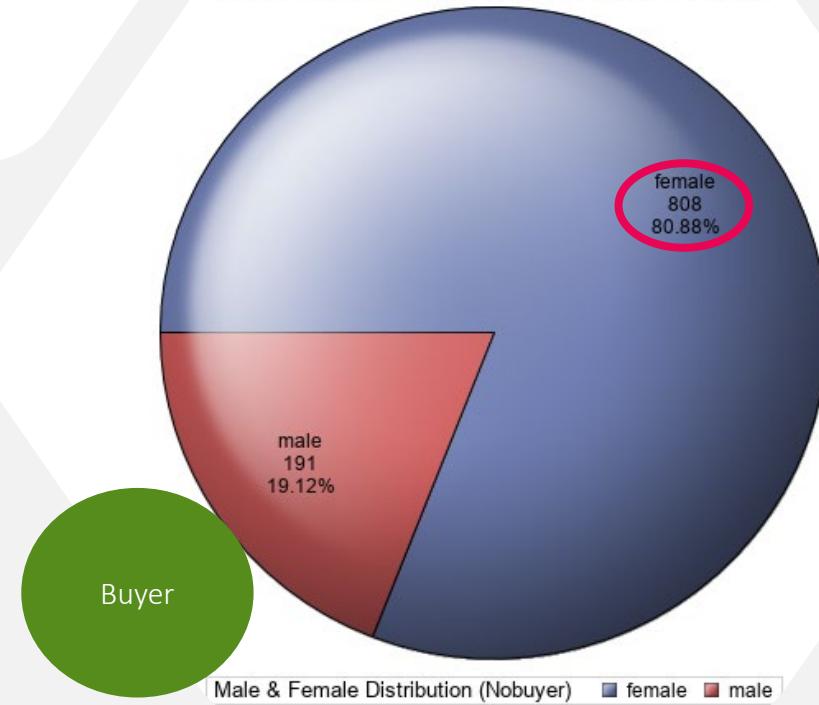
Dataset

Number of Males & Females in Nobuyer Segment

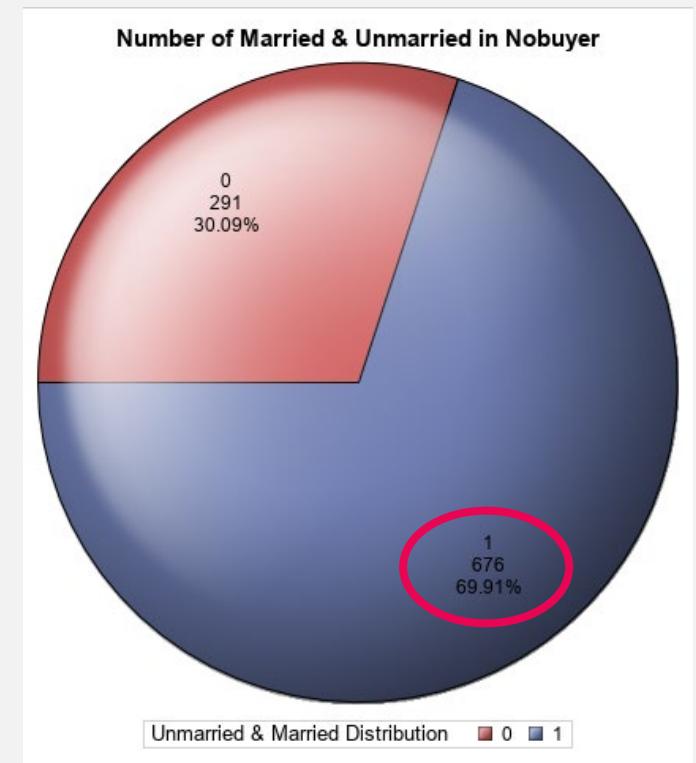
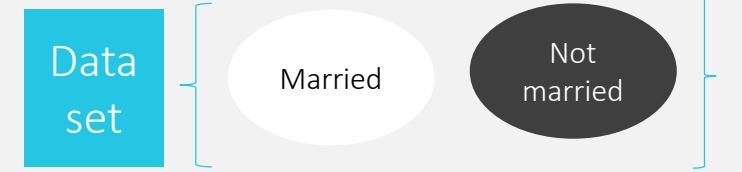
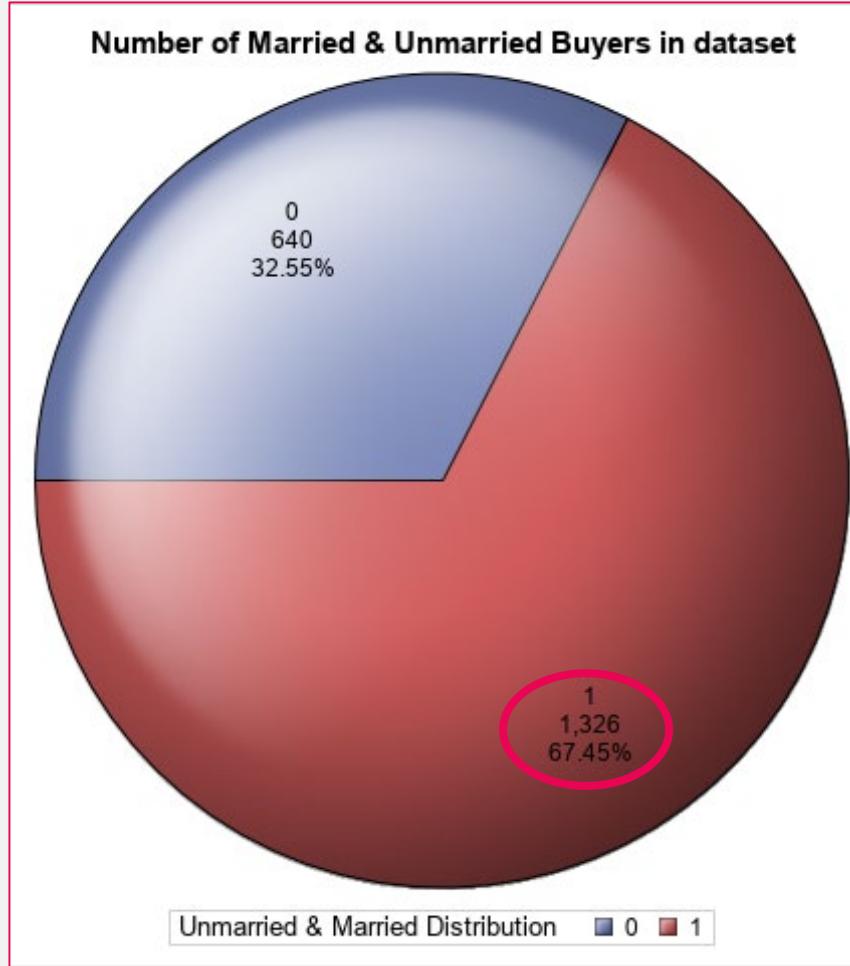
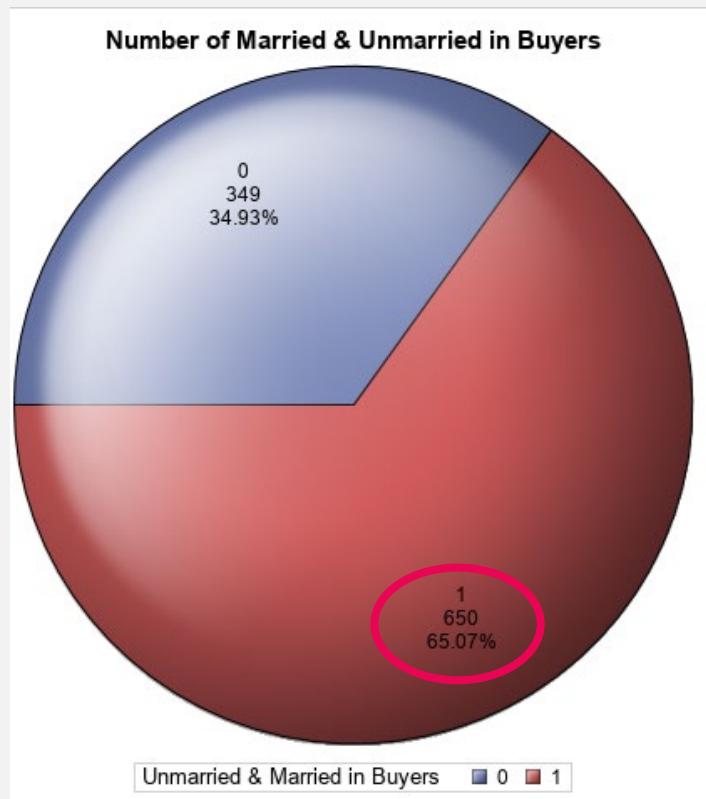
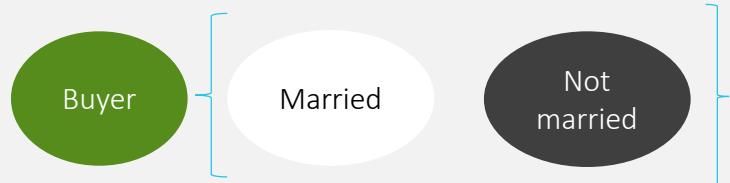


No buyer

Number of Males & Females in Buyer Segment



Buyer-Married buyer unmarried buyer



Buyer & Married: Frequency

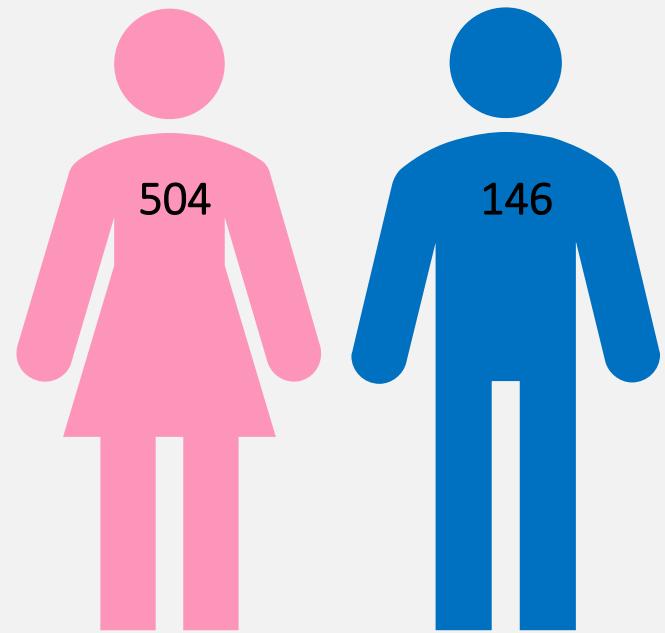
Sex				
SEX	Frequency	Percent	Cumulative Frequency	Cumulative Percent
female	504	77.54	504	77.54
male	146	22.46	650	100.00

Race				
RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	552	84.92	552	84.92
2	70	10.77	622	95.69
4	24	3.69	646	99.38
5	4	0.62	650	100.00

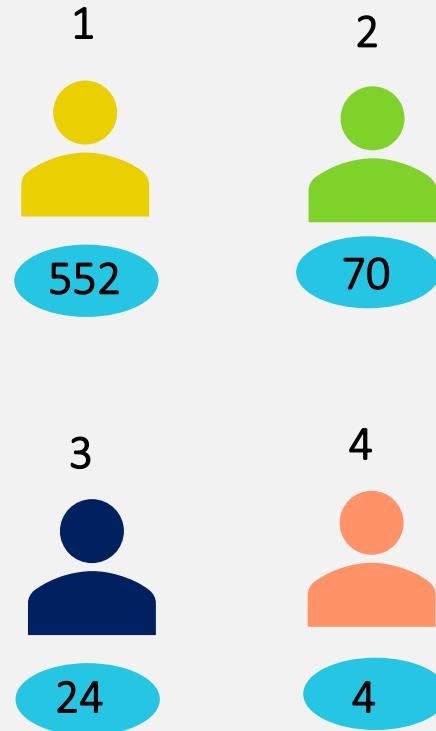
Education Level				
EDLEVEL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	104	16.00	104	16.00
2	241	37.08	345	53.08
3	305	46.92	650	100.00

Buyer & Married : Visualization

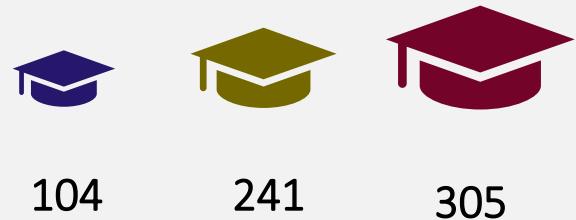
SEX



RACE



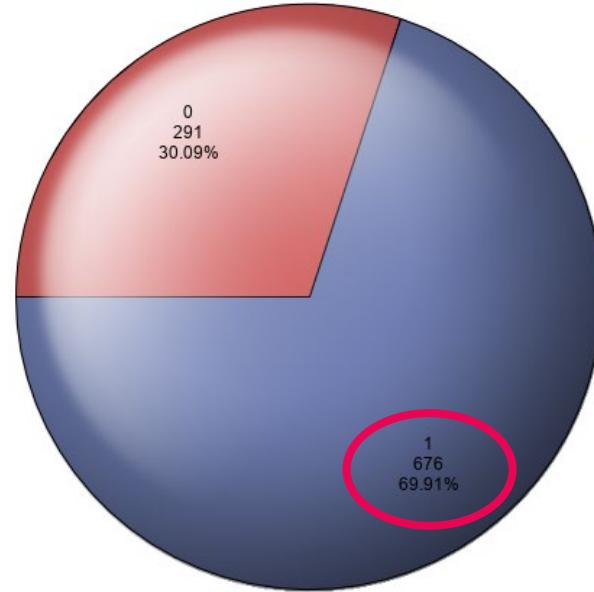
Education Level



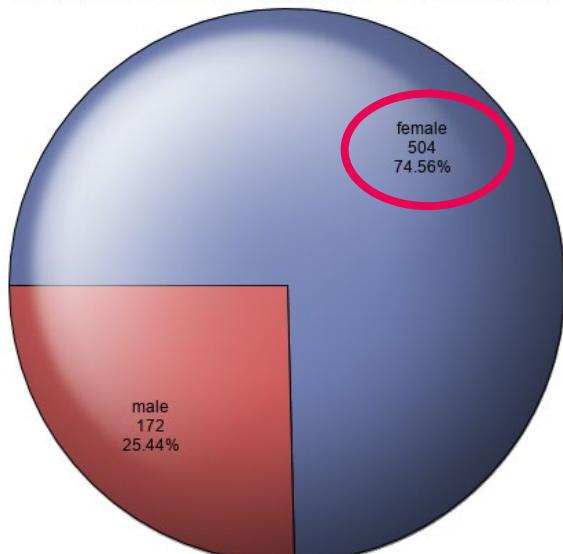
Note: Females, RACE 1 and most educated can be easily targeted.

Who is Not buying?

Number of Married & Unmarried in Nobuyer Segment

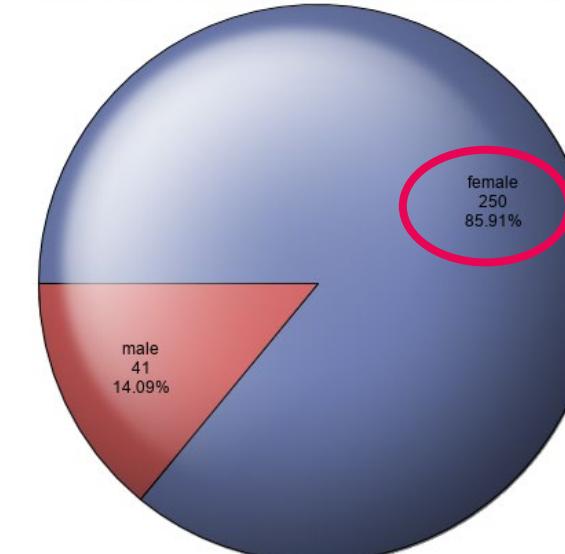


Number of Male & Female in Nobuyer + Married Segment



Male & Female Distribution from Nobuyer & Married blue square = female red square = male

Number of Male & Female in Nobuyer + Unmarried Segment



Male & Female Distribution from Nobuyer & Unmarried blue square = female red square = male

No buyer & Married: Frequency

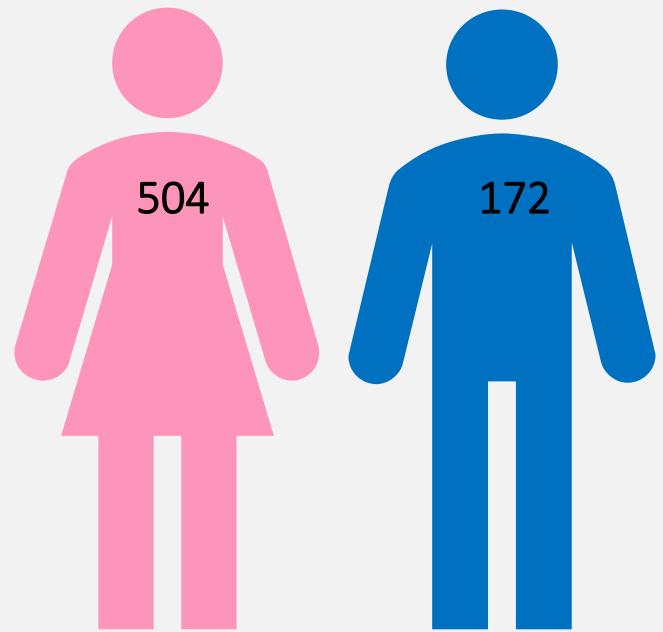
Sex				
SEX	Frequency	Percent	Cumulative Frequency	Cumulative Percent
female	504	74.56	504	74.56
male	172	25.44	676	100.00

Education Level				
EDLEVEL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	97	14.35	97	14.35
2	269	39.79	366	54.14
3	308	45.56	674	99.70
4	2	0.30	676	100.00

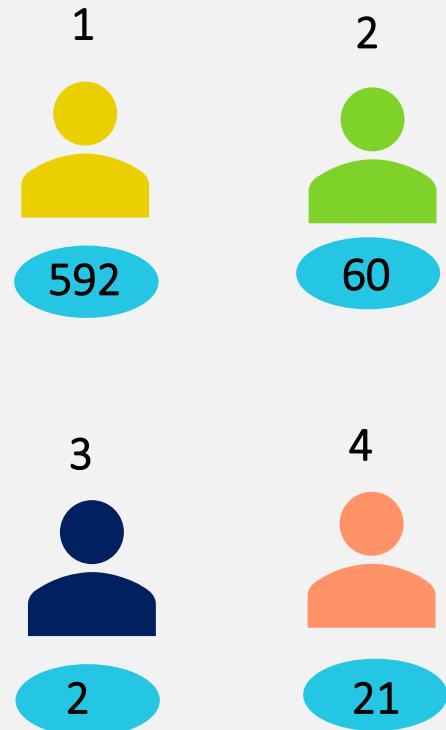
Race				
RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	592	87.57	592	87.57
2	60	8.88	652	96.45
3	2	0.30	654	96.75
4	21	3.11	675	99.85
5	1	0.15	676	100.00

No buyer & Married : Visualization

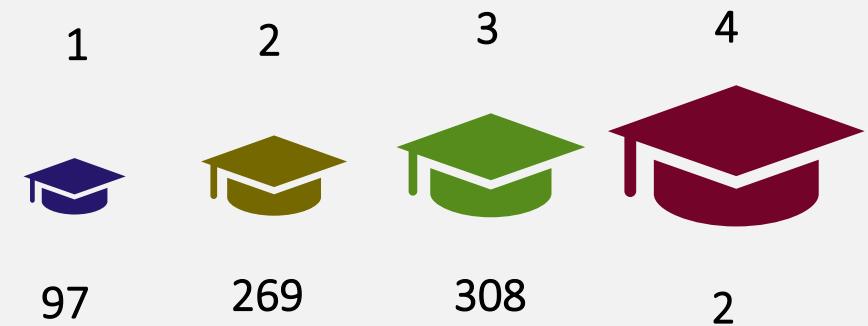
SEX



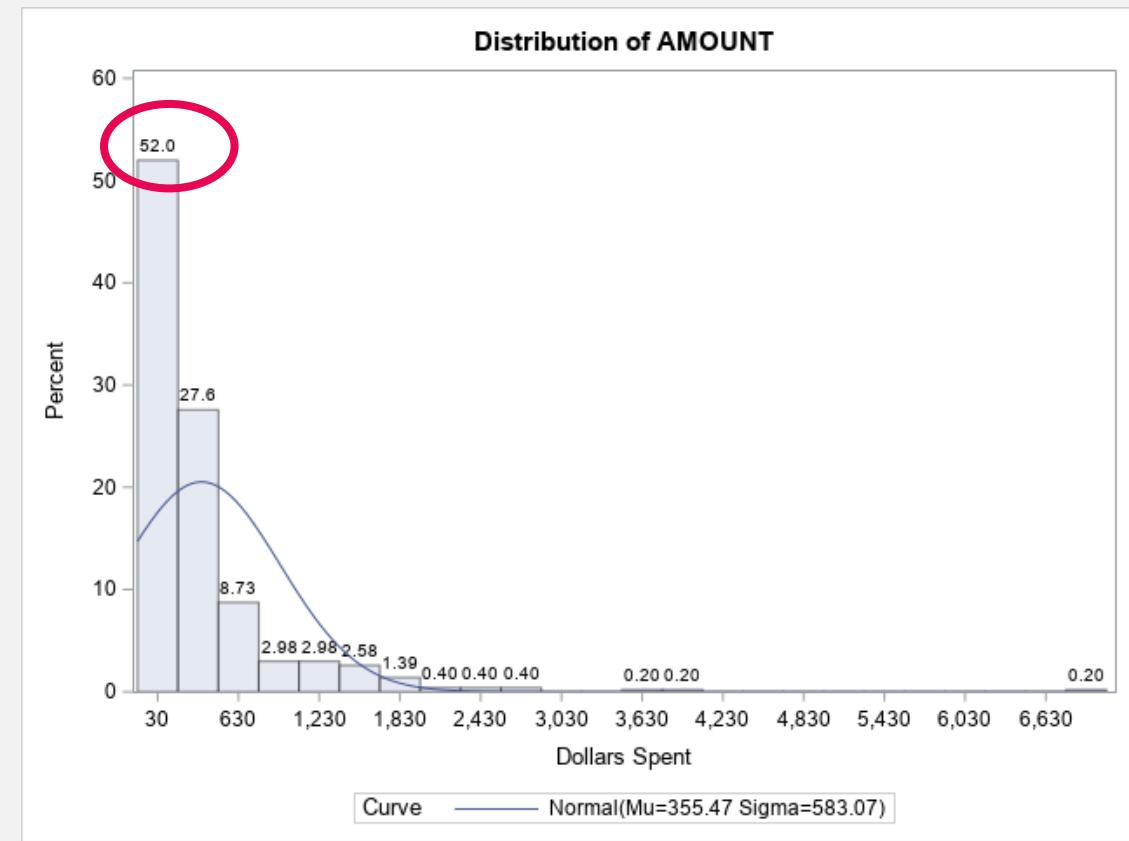
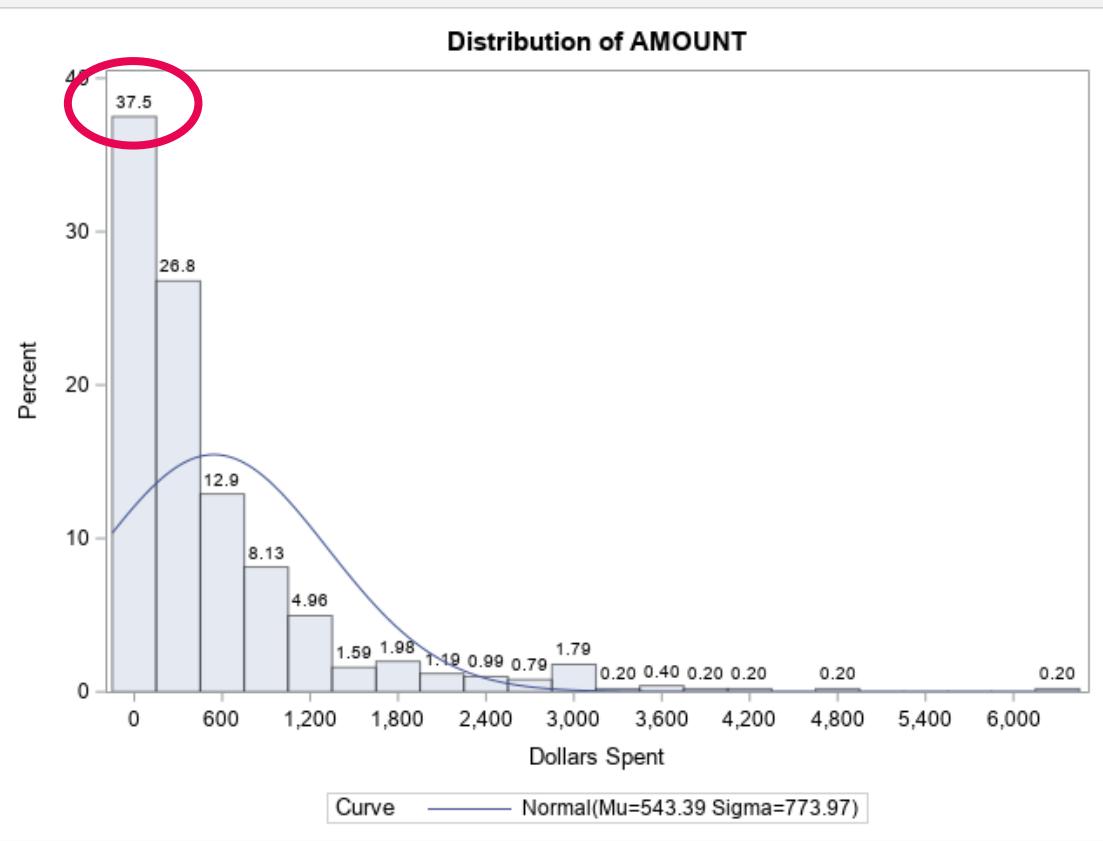
RACE



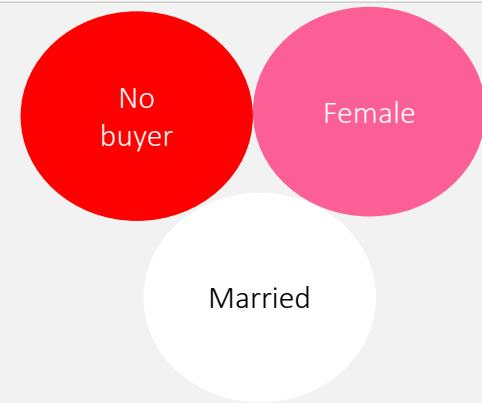
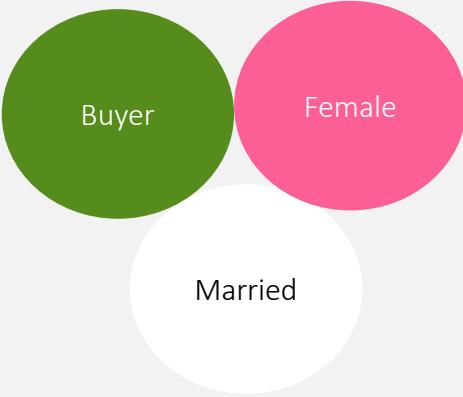
Education Level

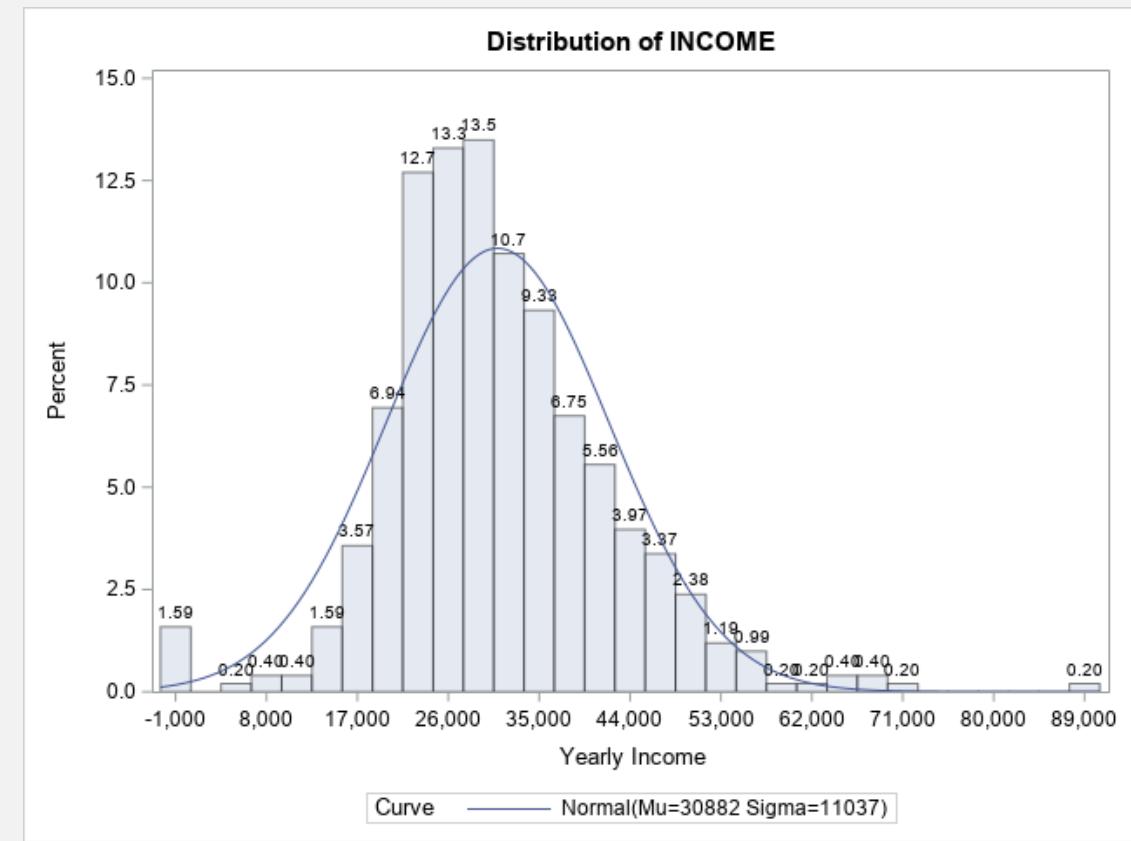
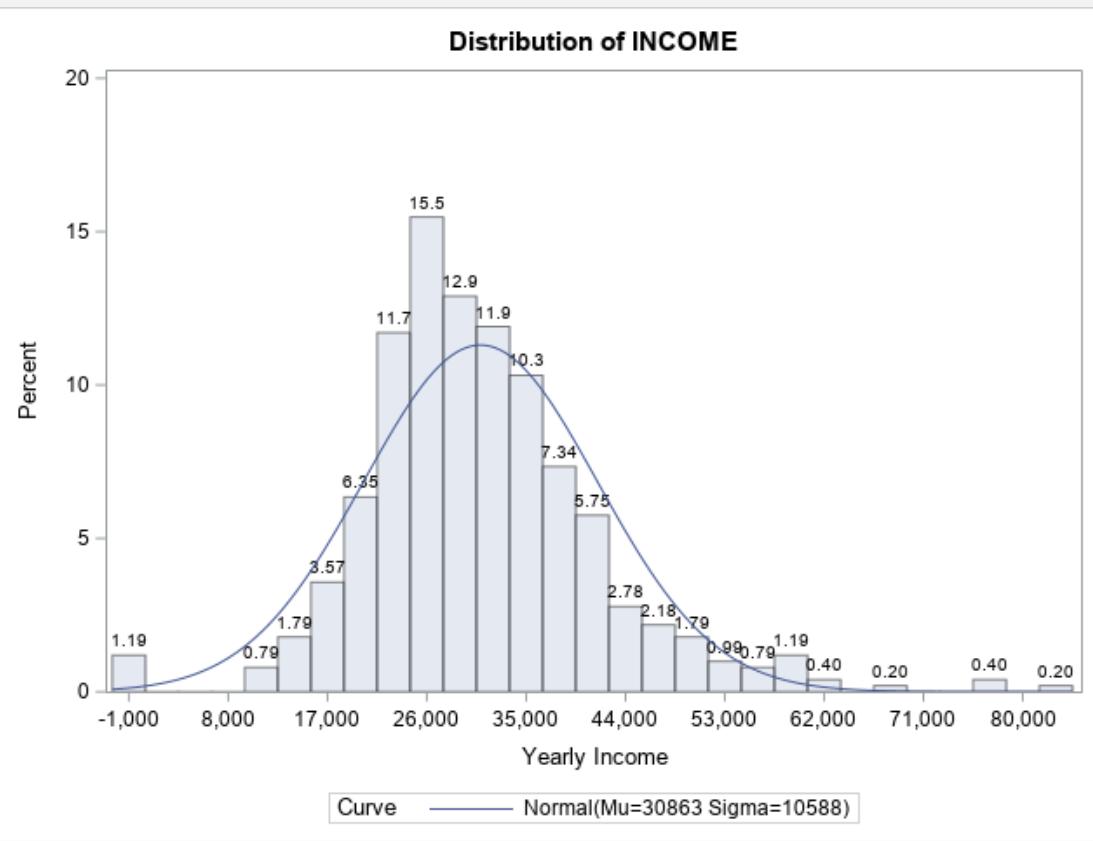


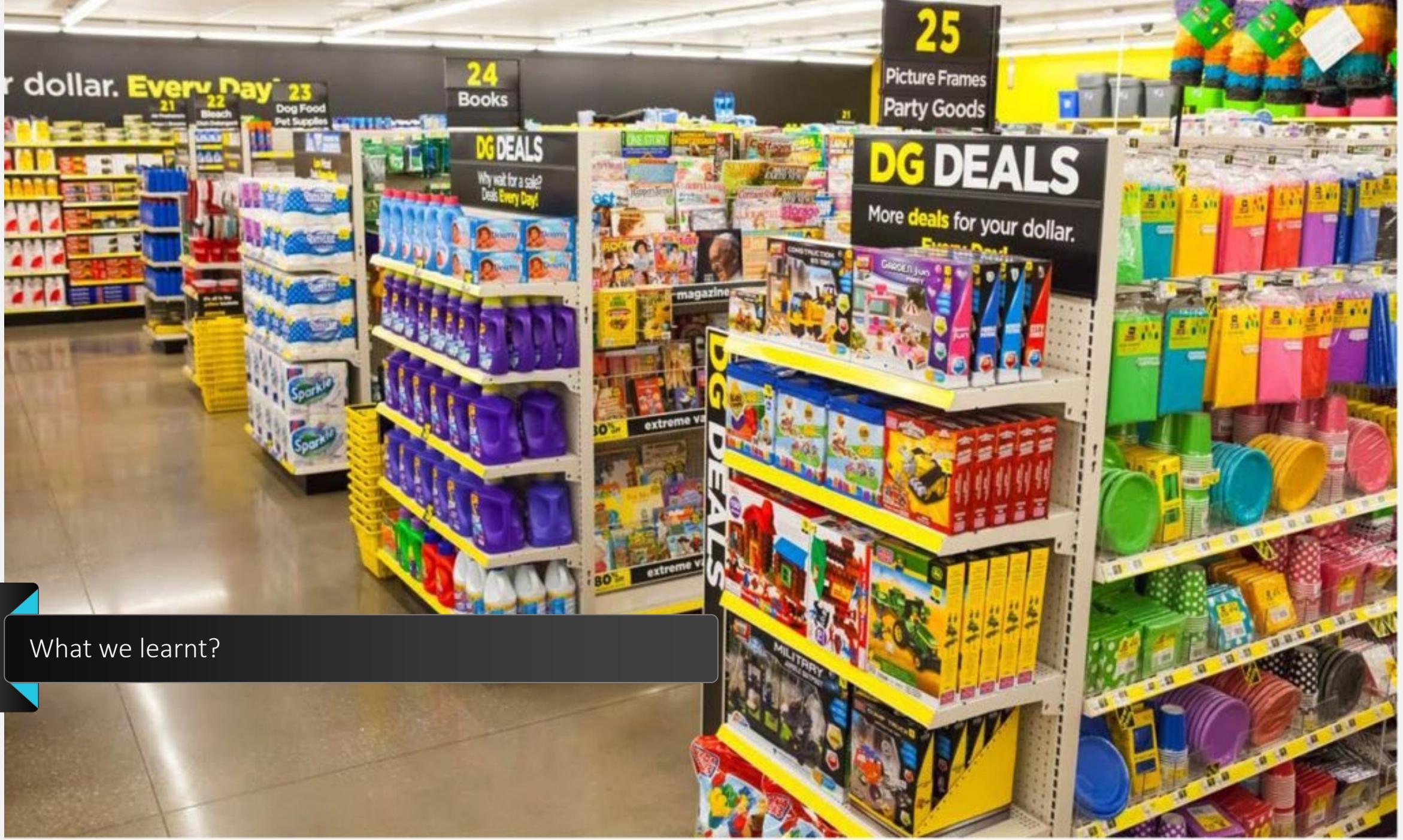
Note: Females, RACE 1 and most educated can be easily targeted.



Why? Difference in Amount Spent







What we learnt?

Takeaways for marketing

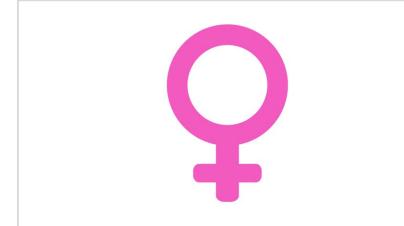


1

Demographics: Marital Status & Gender



Married |
65% buyers are married



Females |
80% buyers are females

2

Demographics: Ethnicity & Education



RACE |
85% buyers are white



Education |
50% buyers have 2nd highest education

3



Income vs Expenditure: Positive Correlation

MORE INCOME → More DOLLARS SPENT |
Married white females with high
education are more likely to buy the
product



Questions?



Annexure

```
* create pie chart for males & females from buyer segment;
```

```
PROC TEMPLATE;
```

```
DEFINE STATGRAPH pie;
```

```
BEGINGRAPH;
```

```
    entrytitle "Number of Males & Females in Buyer Segment";
```

```
LAYOUT REGION;
```

```
    PIECHART CATEGORY = sex /
```

```
    DATALABELLOCATION = INSIDE
```

```
    DATALABELCONTENT = ALL
```

```
    CATEGORYDIRECTION = CLOCKWISE
```

```
    DATASKIN = SHEEN
```

```
    START = 180 NAME = 'pie';
```

```
    DISCRETELEGEND 'pie' /
```

```
    TITLE = 'Male & Female Distribution (Buyer)';
```

```
    ENDLAYOUT;
```

```
ENDGRAPH;
```

```
END;
```

```
RUN;
```

```
PROCSGRENDER DATA = khanproj.buyer
```

```
    TEMPLATE = pie;
```

```
RUN;
```

```
* create pie chart for males & females from nobuyer segment;
```

```
PROC TEMPLATE;
```

```
DEFINE STATGRAPH pie;
```

```
BEGINGRAPH;
```

```
    entrytitle "Number of Males & Females in Nobuyer Segment";
```

```
LAYOUT REGION;
```

```
    PIECHART CATEGORY = sex /
```

```
    DATALABELLOCATION = INSIDE
```

```
    DATALABELCONTENT = ALL
```

```
    CATEGORYDIRECTION = CLOCKWISE
```

```
    DATASKIN = SHEEN
```

```
    START = 180 NAME = 'pie';
```

```
    DISCRETELEGEND 'pie' /
```

```
    TITLE = 'Male & Female Distribution (Buyer)';
```

```
    ENDLAYOUT;
```

```
ENDGRAPH;
```

```
END;
```

```
RUN;
```

```
PROCSGRENDER DATA = khanproj.nobuyer
```

```
    TEMPLATE = pie;
```

```
RUN;
```

```
* create pie chart for married & unmarried from nobuyer segment;
```

```
PROC TEMPLATE;
```

```
DEFINE STATGRAPH pie;
```

```
BEGINGRAPH;
```

```
    entrytitle "Number of Married & Unmarried in dataset";
```

```
LAYOUT REGION;
```

```
    PIECHART CATEGORY = Marital /
```

```
    DATALABELLOCATION = INSIDE
```

```
    DATALABELCONTENT = ALL
```

```
    CATEGORYDIRECTION = CLOCKWISE
```

```
    DATASKIN = SHEEN
```

```
    START = 180 NAME = 'pie';
```

```
    DISCRETELEGEND 'pie' /
```

```
    TITLE = 'Unmarried & Married Distribution';
```

```
    ENDLAYOUT;
```

```
ENDGRAPH;
```

```
END;
```

```
RUN;
```

```
PROCSGRENDER DATA = khanproj.project1
```

```
    TEMPLATE = pie;
```

```
RUN;
```

```
* martial_status variable against purchase (y/m) variable;
```

```
proc freq data = khanproj.marriedbuyer;
tables PURCHASE*MARITAL
/chisq
;
run;
```

```
* Getting a snapshot about the data;
```

```
ODS pdf FILE = "C:\users\usman\desktop\sas_proj\data_contents.pdf";
```

```
PROC CONTENTS DATA = khanproj.project1;
RUN;
ODS pdf close;
```

```
* Separating the femaleonly;
```

```
PROC SQL;
```

```
CREATE table femaleonly AS
select * from khanproj.project1 where SEX='female';
quit;
```

```
* Separating the maleonly ;
```

```
PROC SQL;
```

```
CREATE table maleonly AS
select * from khanproj.project1 where SEX='male';
quit;
```

```
* Separating the marriedbuyer;
```

```
PROC SQL;
```

```
CREATE table marriedbuyer AS
select AMOUNT, INCOME, HOMEVAL, AGE, EDLEVEL, RACE, SEX from
khanproj.project1 where marital=1 AND PURCHASE=1;
quit;
```

```
proc univariate data = khanproj.marriedbuyer noprint;
```

```
histogram INCOME
```

```
/
```

```
normal (
```

```
mu = est
```

```
sigma = est
```

```
color = blue
```

```
w = 2.5
```

```
)
```

```
barlabel = percent
```

```
midpoints = 6000 to 12000 by 4000;
```

```
run;
```

```
proc univariate data = khanproj.nomarriedbuyer noprint;
```

```
histogram INCOME
```

```
/
```

```
normal (
```

```
mu = est
```

```
sigma = est
```

```
color = blue
```

```
w = 2.5
```

```
)
```

```
barlabel = percent
```

```
midpoints = 6000 to 12000 by 4000;
```

```
run;
```

```
PROC UNIVARIATE DATA = marriedbuyer;
```

```
VAR INCOME;
```

```
RUN;
```

```
PROC SGPANEL DATA = khanproj.nomarriedbuyer;
```

```
PANELBY PURCHASE/ columns = 1 novarnname;
```

```
VBOX AGE / category = RACE;
```

```
PROC UNIVARIATE DATA = nomarriedbuyer;
```

```
VAR INCOME;
```

```
RUN;
```

```
title 'Horsepower of cars by types';
```

```
RUN;
```

```
proc FREQ data = khanproj.nobuyermarried ;
```

```
tables SEX RACE EDLEVEL ;
```

```
run;
```

```
PROC UNIVARIATE DATA = khanproj.nobuyerfemale;
```

```
VAR INCOME;
```

```
RUN;
```

```
PROC SGPANEL DATA = khanproj.marriednobuyer;
```

```
PANELBY PURCHASE/ columns = 1 novarnname;
```

```
VBOX JOB / category = RACE;
```

```
proc FREQ data = buyermale ;
```

```
tables JOB RACE ;
```

```
run;
```

```
PROC UNIVARIATE DATA = khanproj.buyermale;
```

```
VAR INCOME;
```

```
RUN;
```

```
title 'Horsepower of cars by types';
```

```
RUN;
```

```
proc FREQ data = buyerfemale ;
```

```
tables JOB RACE ;
```

```
run;
```

```
PROC UNIVARIATE DATA = khanproj.nobuyermale;
```

```
VAR INCOME;
```

```
RUN;
```

```
PROC SGPANEL DATA = khanproj.nomarriednobuyer;
```

```
PANELBY PURCHASE/ columns = 1 novarnname;
```

```
VBOX JOB / category = RACE;
```

```
proc reg data = nobuyermale;
```

```
model AMOUNT = INCOME, HOMEVAL ;
```

```
run;
```

```
title 'Horsepower of cars by types';
```

```
RUN;
```

```
proc corr data = khanproj.nobuyermale rank plots(only) = scatter(elipse=none);
```

```
VAR INCOME HOMEVAL;
```

```
with AMOUNT;
```

```
title 'Age of Account vs Other Selected Variables';
```

```
run;
```

That's all folks!

Buyer & Unmarried: Frequency

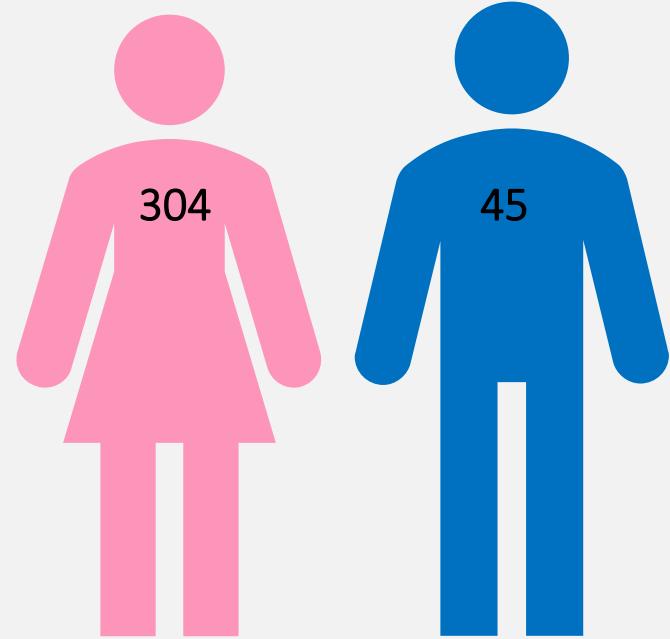
Sex				
SEX	Frequency	Percent	Cumulative Frequency	Cumulative Percent
female	304	87.11	304	87.11
male	45	12.89	349	100.00

Race				
RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	268	76.79	268	76.79
2	62	17.77	330	94.56
4	15	4.30	345	98.85
5	4	1.15	349	100.00

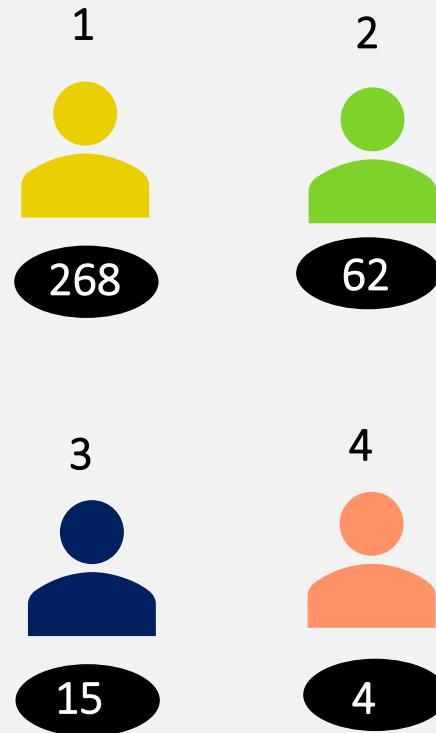
Education Level				
EDLEVEL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	43	12.32	43	12.32
2	124	35.53	167	47.85
3	180	51.58	347	99.43
4	2	0.57	349	100.00

Buyer & Unmarried : Visualization

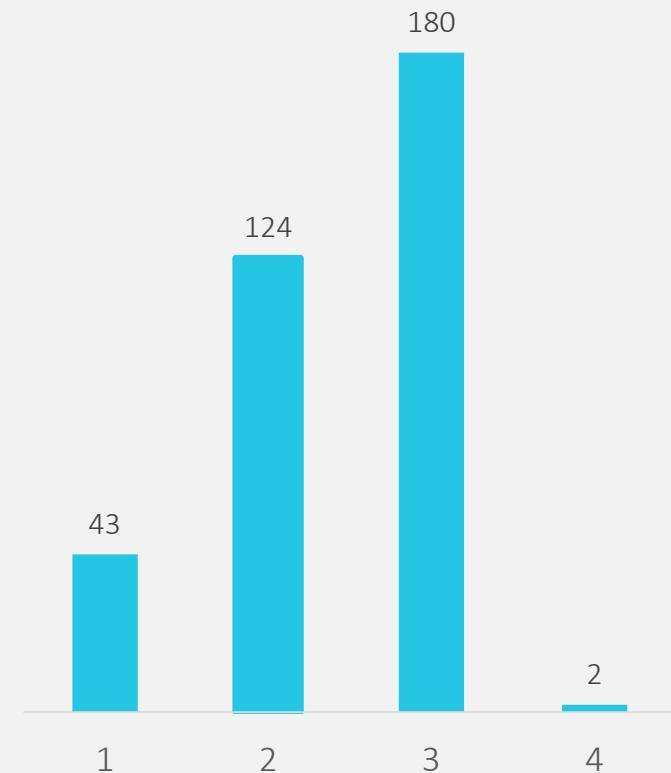
SEX



RACE



Education Level



Note: Females, RACE 1 and most educated can be easily targeted.

Framework

The selected features for analyzing purchase behavior

Yearly Income

Correlation

Home value

Correlation

Married

Chi Square

Race

Chi Square

Dollars Spent (\$)

Age

Education Level

SEX

Purchase (Y/N)

