Group B2

Aman Rathi (61310645)
Arpita Ray (61310120)
Akshat Narain (61310333)
Jayagowri Hoblidar (61310353)
Vismay Shah (61310258)

EFFECTIVE PREMIUM

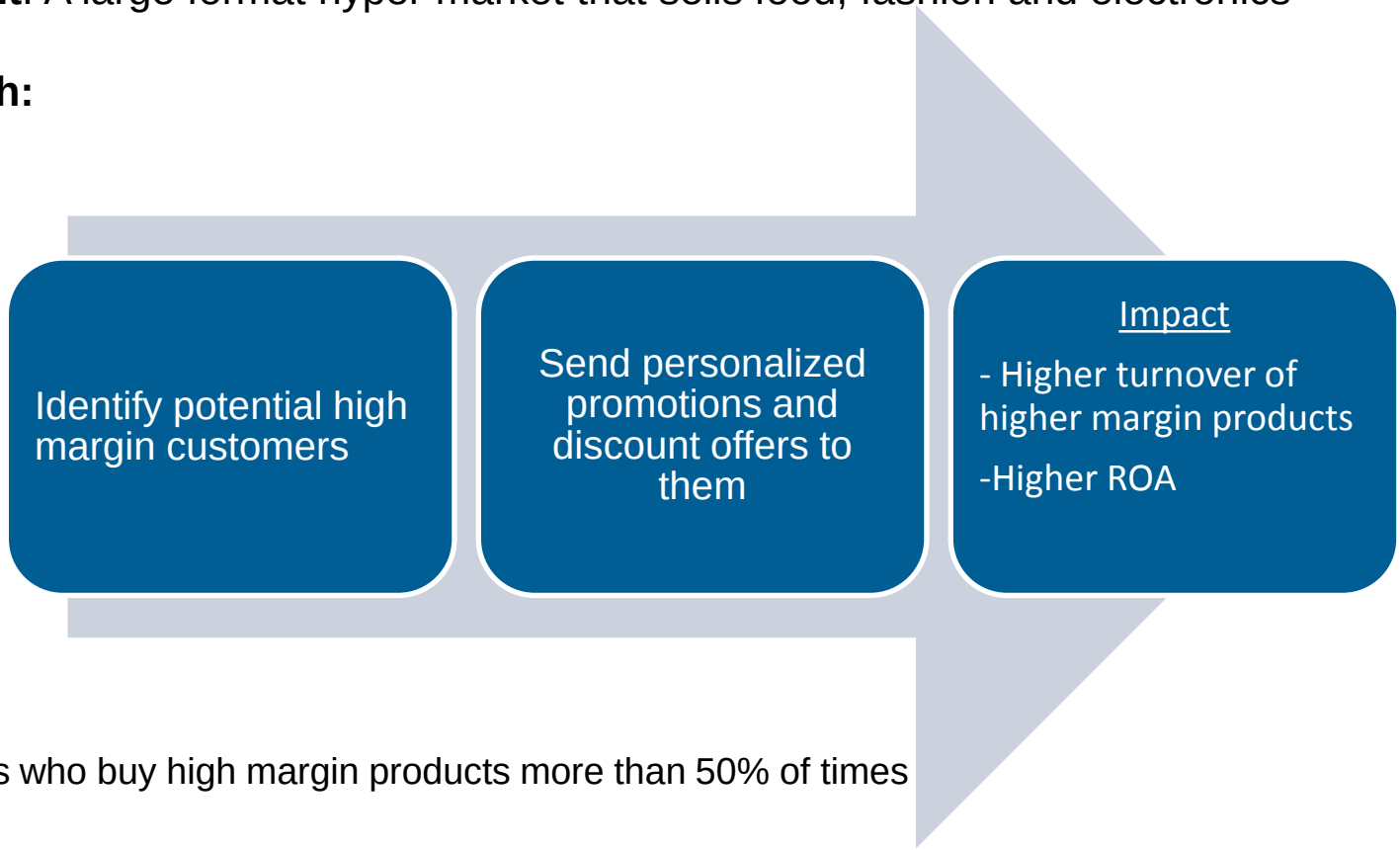
CUSTOMER TARGETING USING CLASSIFICATION METHODS

Business Goal

Business Goal: Increase sales of high margin products at Hypermart by effectively targeting high margin customers*
(*thereby increasing the sales per square footage of stores and Return on Assets)*

Our client: A large format hyper market that sells food, fashion and electronics

Approach:



*customers who buy high margin products more than 50% of times

Data Mining Goal

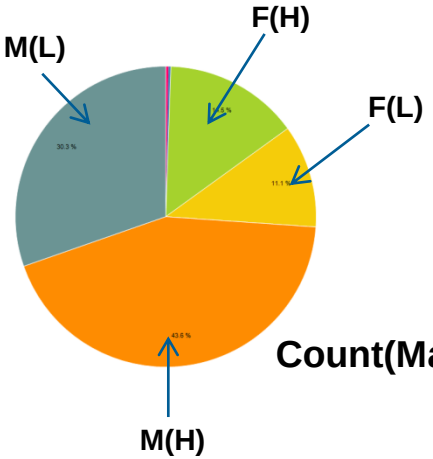
- To identify the potential high margin customers based on their demographics and purchasing information on their first purchase.

- A supervised learning method would be used with predictors and categorical outcome variable

- **Predictors** initially include all columns from the transactions and customer data files. For e.g. demographical columns such as age, marital status etc., and transactional data such as quantity sold, price of basket etc. We use **best subset selection** from the tools to select the best predictors later.

- **Outcome variable** is **categorical** which indicates whether the customer is a potential buyer of high margin or low margin products

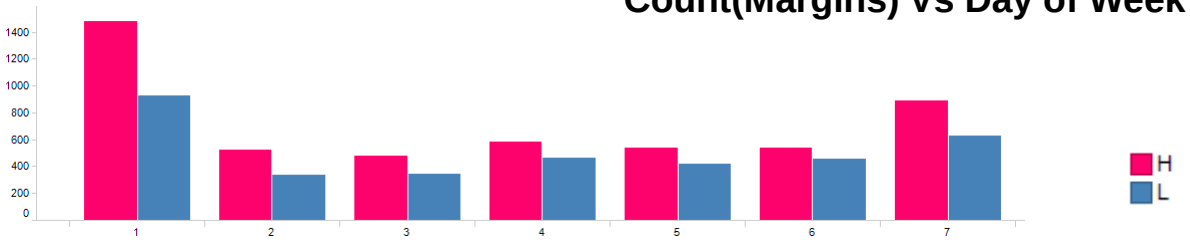
Data Exploration & Visualization



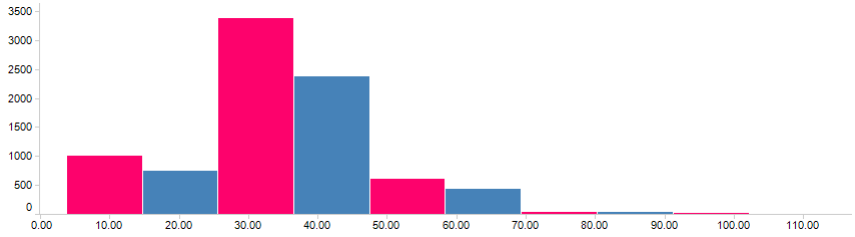
Count(Margins) Vs Sex

1. Transaction data was merged with the customer dataset. Each transaction was classified as high or low margin using secondary dataset.
2. Rolled this up into customer level and identified each customer as High or Low margin customer depending on cutoff of 0.5
3. Filtered out the customers who had single basket data. Of the remaining retained only the first basket data for all customers

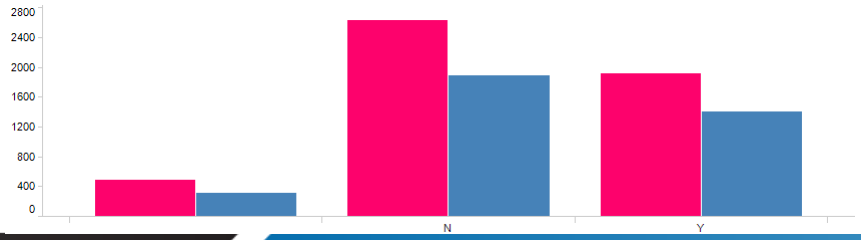
Count(Margins) Vs Day of Week



Count(Margins) Vs Age



Count(Margins) Vs Marital Status



Models & Benchmarking

Benchmark : An error % less than 40% (Naïve Rule error for H→ L)

Logistic Regression

Best !

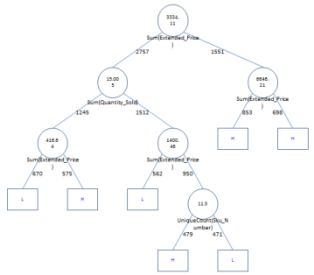
#Coeffs	RSS	Cp
9	4298.7832	9.7851095

Classification Confusion Matrix		
	Predicted Class	
Actual Class	H	L
H	826	203
L	422	272

Error Report			
Class	# Cases	# Errors	% Error
H	1029	203	19.73
L	694	422	60.81
Overall	1723	625	36.27

Predictors (Best Subset) : # of subdepts, Quantity Sold, Price of Basket, Age, Sex, Day of week

CART



Classification Confusion Matrix		
	Predicted Class	
Actual Class	H	L
H	736	293
L	342	352

Error Report			
Class	# Cases	# Errors	% Error
H	1029	293	28.47
L	694	342	49.28
Overall	1723	635	36.85

Ensembles (Logistic Regression + CART) Results

H→L (Error : 21.19%)

Recommendations & Insights

- **Recommendations**
  - ✓ Focus on the **following groups** as they are more likely to be high margin customers
    - Males (3 times more likely)
    - Age group b/w 25-45 (3 times more likely)
    - Unmarried (1.3 times more likely)
  - ✓ Execute the model in **real time** as the customer checks out, and give coupons if he/she is a high margin first time buyer.
  - ✓ Don't let the model become stale. Continue to **collect data periodically** to refine the model in future

- **Improving the Model**
  - ✓ Improving the prediction by collecting **better predictors** like income etc.
  - ✓ Getting **numbers on costs** for incorrectly classifications to develop accurate confusion matrices

Thank You

Q & A