CAPSTONE PROJECT ON OFFICE SUPPLY STORE

BY: HAFTOM ABRHA

OFFICE SUPPLY STORE

Contents

Background

- An Office Supply Store tests a telemarketing campaign for its existing customers(about 16,000)
- The client considers a prior campaign response data and likes it as input for maximizing future campaign potentials:
  - Develop model to understand targeting
  - Minimize cost to maximize profit by not contacting customers randomly
- Also to target future campaigns on those likely to respond to the campaign.
  - Develop models
  - Understand targeting
  - Targeting all customers would cost $1.4M against the 100k customer base
- The store's response data snapshots:
  - Historical sales
  - Prior year number of transactions
  - List of targeted customers:
    - # of Employees
    - Language spoken
    - Prior products purchased
    - Repurchase Method
  - Date of first purchase

Objectives

- Analyze response data set to optimize future campaign targeting and revenue
- Provide customer prioritization based on profitability

Methodology

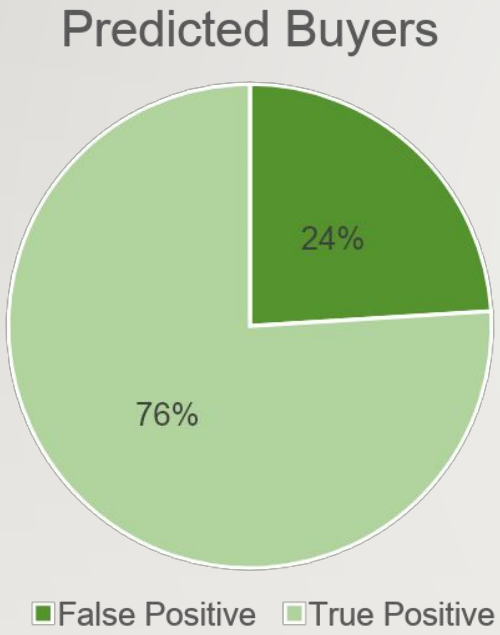
1. Perform EDA on dataset
2. Interview client to determine use case
3. Transform dataset
   - Bucketing
   - Imputing
   - Feature engineering
4. Perform data model creation
   - Logistic Regression to determine probability of response
   - Linear Regression to determine likely size of purchase
   - Calculate profitability based on regression outputs, gross margin, and campaign and transaction cost
5. Define customer deciles
6. Create lift table
7. Deliver recommendation

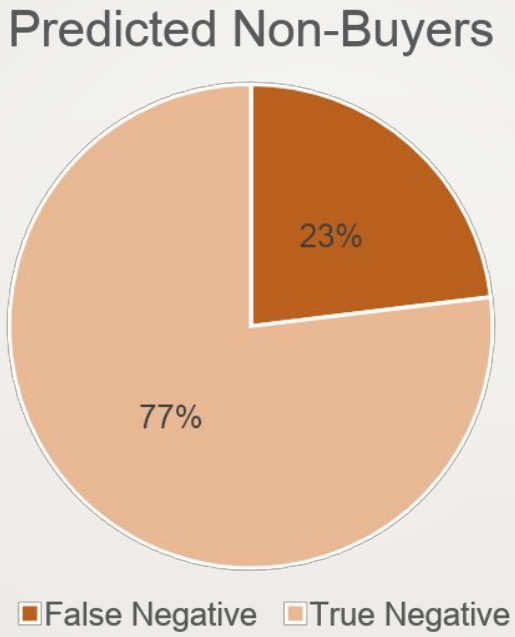
EDA

- Some duplicate customer ID's
- Removed outliers in # of prior year transactions
- Created binary variable for 'Has Campaign Sales'
- Feature engineering:
  - Tenure
  - All product variables changed to binary
  - Total type (product purchase mix)
  - Communication methods
- Consolidation
  - Last transaction channel
    - Split into 'branch' and 'other'
  - Campaign sales & historical sales volume
    - Zeroed negatives
  - Tenure
    - Bucketed into tertiles
  - Sales
    - Bucketed into quartiles

Stepwise Logistic Regression

- Independent Validation Sample Size 8,089





- Significant Variables:
  - High Historical Sales Volume
  - Large # of Prior Year Transactions
  - Long Tenure
  - Purchased Computer
  - Purchased Monitor
  - Purchased Office Supplies
  - Purchased Printer
  - Non Auto-Renew or Notice Repurchase Method
  - Purchased Standard Chair

Stepwise Linear Regression

- Significant Variables: 
  - Greater # of Prior Year Transactions
  - Purchased Monitor
  - Greater # of Employees
  - Purchased Office Supplies
  - Non-Branch Last Transaction Channel
  - High Historical Sales Volume
  - Long Tenure

Profitability Model

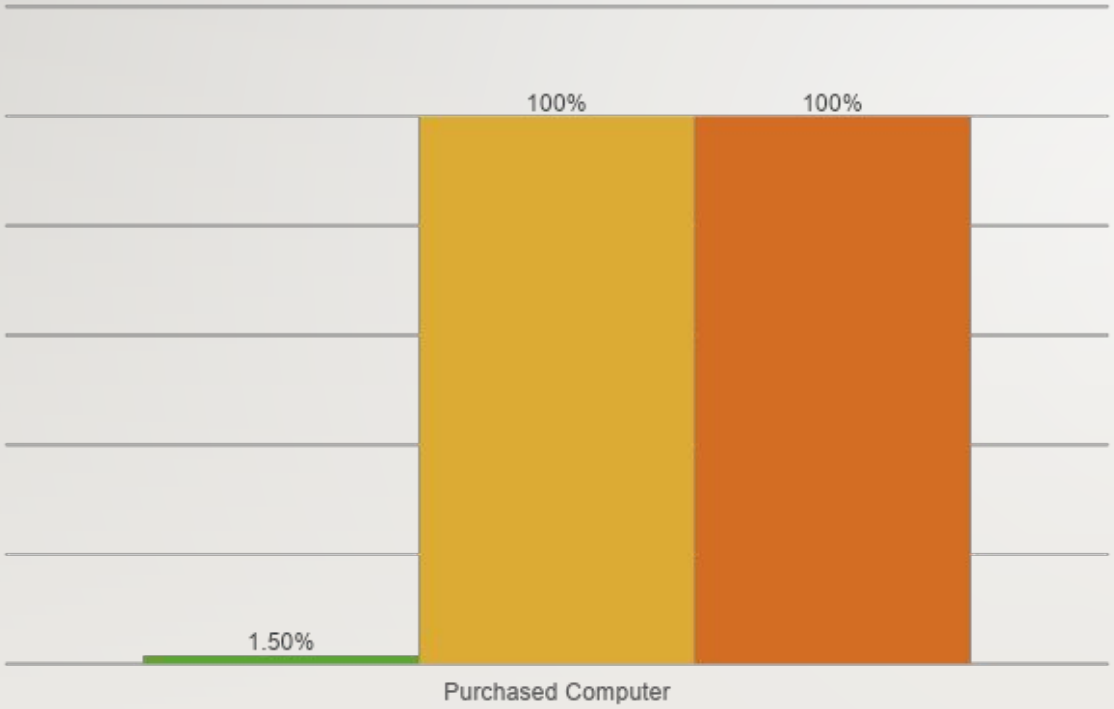
Gross Margin on Sales: 22%
Campaign Cost: $45.65 per business contacted
Transaction Cost: $8.40 per transaction

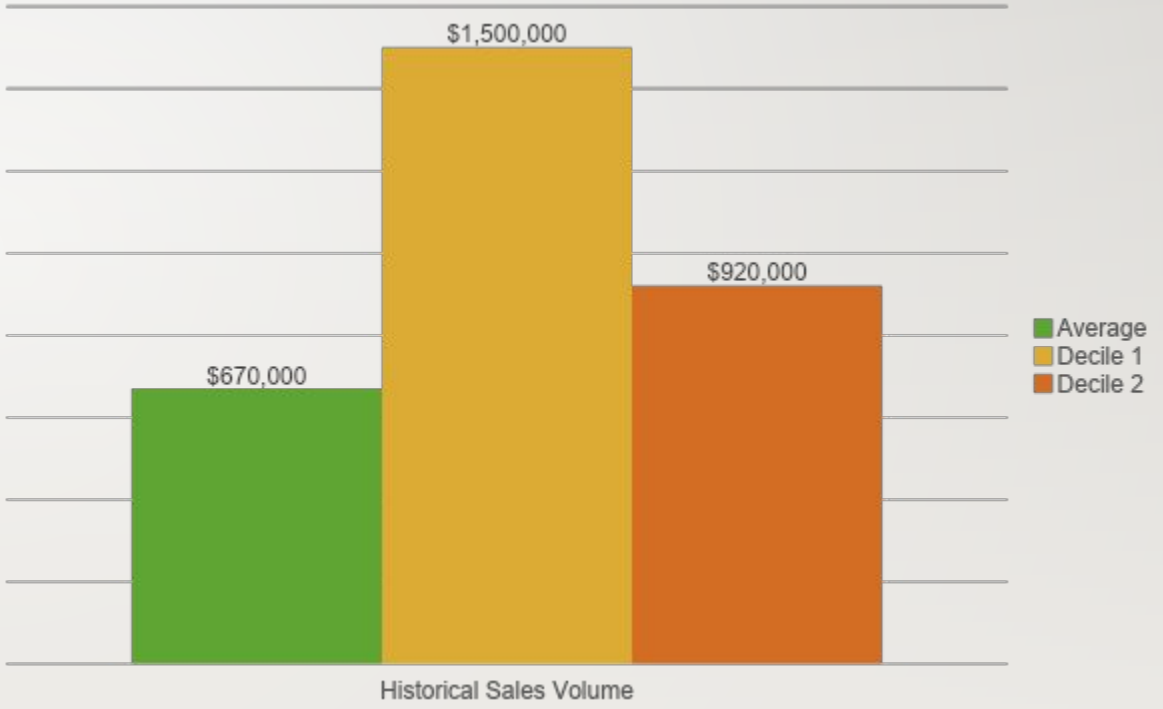
$$E(Profit) = .22 * Prob(Sale) * Est(Transaction\ Size) - \$8.40 * Prob(Sale) - \$45.65$$

Lift

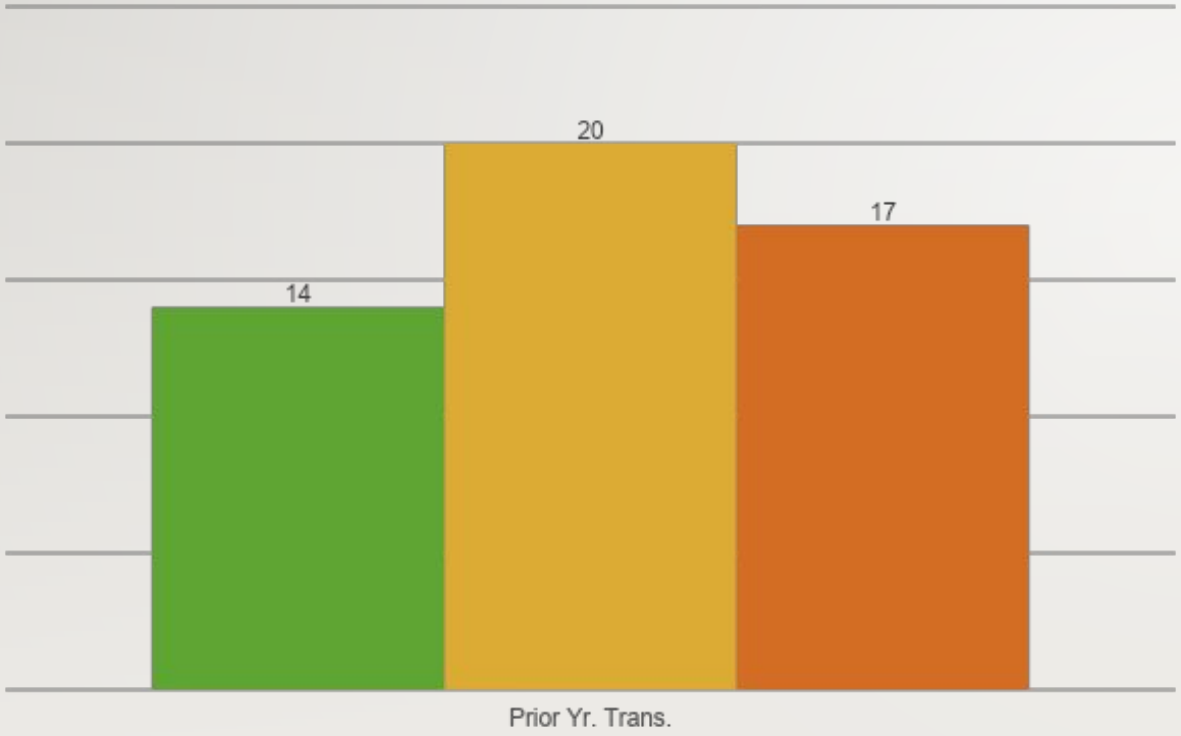
Decile	Number of Customers	Actual Profitabilty Per Customer	Lift Over Average	Total Profit	Incr Proj Profit 100k Cust Base ($K)	Total Proj Profit 100k Cust Base ($K)	Cuml Incr Profit 100k Cust Base ($K)	Cuml Total Profit 100k Cust Base ($K)
1	113	$ 63	$ 77	$ 7,062.50	$ 766	$ 625	$ 766	$ 625
2	111	$ 8	$ 22	$ 877	$ 220	$ 79	$ 986	$ 704
3	109	$ (8)	$ 6	$ (872)	$ 61	$ (80)	$ 1,047	$ 624
4	113	$ (16)	$ (2)	$ (1,842)	$ (22)	$ (163)	$ 1,025	$ 461
5	107	$ (22)	$ (8)	$ (2,311)	$ (75)	$ (216)	$ 950	$ 245
6	112	$ (25)	$ (11)	$ (2,845)	$ (113)	$ (254)	$ 837	$ (9)
7	113	$ (29)	$ (15)	$ (3,322)	$ (153)	$ (294)	$ 684	$ (303)
8	110	$ (33)	$ (19)	$ (3,630)	$ (189)	$ (330)	$ 495	$ (633)
9	118	$ (37)	$ (23)	$ (4,354)	$ (228)	$ (369)	$ 267	$ (1,002)
10	104	$ (41)	$ (27)	$ (4,264)	$ (269)	$ (410)	$ (2)	$ (1,412)
Total	1,110	$ (14.1)	$ (0)	$ (15,501)	$ (0)	$ (141)	$ (2)	

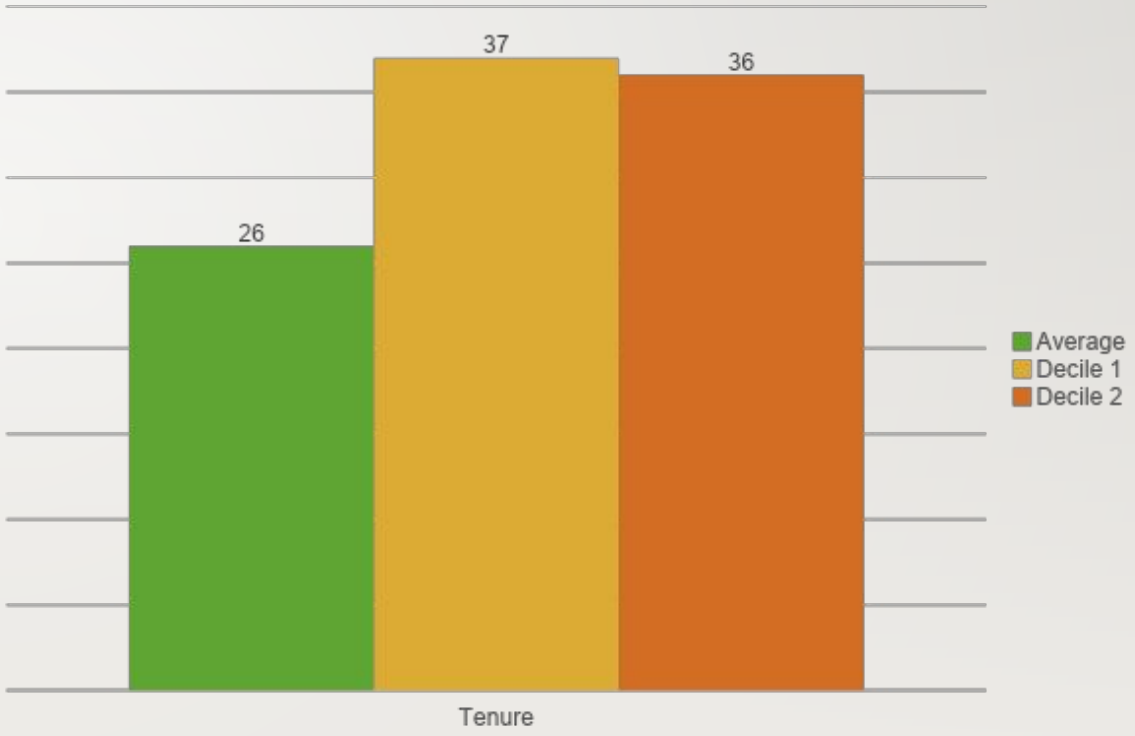
Decile Profile





Decile Profile Cont.

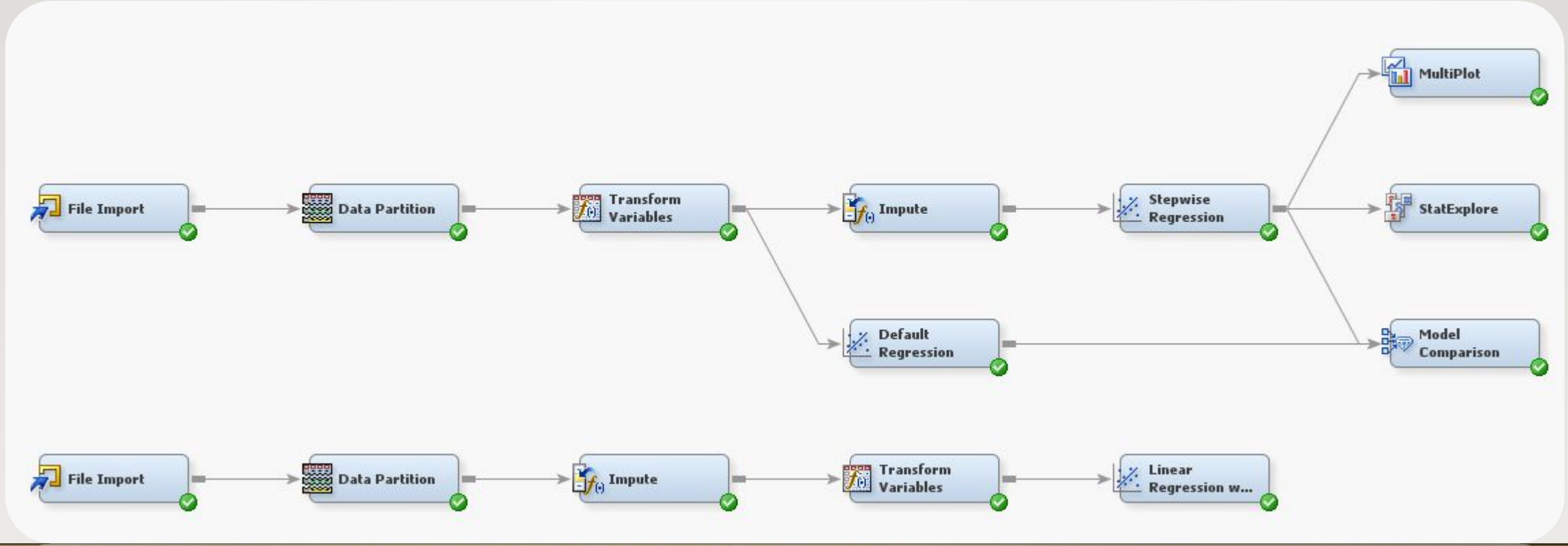




Recommendations

- Target customers in deciles 1 & 2
  - Potential profit impact: $700k
- Low cost targeted email campaign for potential buyers in deciles 3 & 4
- Customer profile refinement
  - What other information can we gather about our base to reduce false negatives and missed sales?
- Database updates to introduce new variable groupings
- Hire analyst to maintain and improve model

Appendix 1 - Model



Appendix 2 - Analysis Results

AVERAGE

- 26 Year Tenure
- $250 in campaign sales
- $670k in historical sales volume
- 1.5% Purchased Computer
- Avg. of prior year transactions: 14

CANDIDATE DECILE ONE

- 37 Year Tenure
- $2k in campaign sales
- $1.5M in historical sales volume
- 100% Purchased Computer
- Avg. of prior year transactions: 20

CANDIDATE DECILE TWO

- 36 Year Tenure
- $1.5k in campaign sales
- $920k in historical sales volume
- 100% Purchased Computer
- Avg. of prior year transactions: 17