“Analysis of Marketing Campaign”

Our company?

Large customer base with varying needs

Operates across Store, Catalog, and Online channels

Increasing pressure to personalize marketing efforts

Diverse product categories

Data Set Overview

- It includes 29 variables and 2,240 customer records, making it suitable
  for exploratory data analysis and clustering
- Key Data Components
- Demographics: Income, Age, Family size (Kids/Teens)
- Purchasing Behavior: Product category spend
- Promotion Behavior: Deals, Campaign responses
- Channel Usage: Web, Catalog, Store
- Engagement: Recency, Complaints, Web Visits
-

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
1	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Recency	MntWines	MntFruits	MntMeatPro	MntFishPro	MntSweetP	MntGoldPro	NumDealsP	NumWebPu	NumCatalo	NumStoreP	NumWebVi	AcceptedC	AcceptedC	AcceptedC	AcceptedC	AcceptedC	Complain	Z_CostCon	Z_Re
2	5524	1957	Graduation	Single	58138	0	0	58	635	88	546	172	88	88	3	8	10	4	7	0	0	0	0	0	0	3	
3	2174	1954	Graduation	Single	46344	1	1	38	11	1	6	2	1	6	2	1	1	2	5	0	0	0	0	0	0	3	
4	4141	1965	Graduation	Married	71613	0	0	26	426	49	127	111	21	42	1	8	2	10	4	0	0	0	0	0	0	3	
5	6182	1984	Graduation	Married	26646	1	0	26	11	4	20	10	3	5	2	2	0	4	6	0	0	0	0	0	0	3	
6	5324	1981	PhD	Married	58293	1	0	94	173	43	118	46	27	15	5	5	3	6	5	0	0	0	0	0	0	3	
7	7446	1967	Master	Married	62513	0	1	16	520	42	98	0	42	14	2	6	4	10	6	0	0	0	0	0	0	3	
8	965	1971	Graduation	Divorced	55635	0	1	34	235	65	164	50	49	27	4	7	3	7	6	0	0	0	0	0	0	3	
9	6177	1985	PhD	Married	33454	1	0	32	76	10	56	3	1	23	2	4	0	4	8	0	0	0	0	0	0	3	
10	4855	1974	PhD	Married	30351	1	0	19	14	0	24	3	3	2	1	3	0	2	9	0	0	0	0	0	0	3	
11	5899	1950	PhD	Married	5648	1	1	68	28	0	6	1	1	13	1	1	0	0	20	1	0	0	0	0	0	3	
12	1994	1983	Graduation	Married	51381.5	1	0	11	5	5	6	0	2	1	1	1	0	2	7	0	0	0	0	0	0	3	
13	387	1976	Basic	Married	7500	0	0	59	6	16	11	11	1	16	1	2	0	3	8	0	0	0	0	0	0	3	
14	2125	1959	Graduation	Divorced	63033	0	0	82	194	61	480	225	112	30	1	3	4	8	2	0	0	0	0	0	0	3	
15	8180	1952	Master	Divorced	59354	1	1	53	233	2	53	3	5	14	3	6	1	5	6	0	0	0	0	0	0	3	
16	2569	1987	Graduation	Married	17323	0	0	38	3	14	17	6	1	5	1	1	0	3	8	0	0	0	0	0	0	3	
17	2114	1946	PhD	Single	82800	0	0	23	1006	22	115	59	68	45	1	7	6	12	3	0	0	1	1	0	0	3	
18	9736	1980	Graduation	Married	41850	1	1	51	53	5	19	2	13	4	3	3	0	3	8	0	0	0	0	0	0	3	
19	4939	1946	Graduation	Married	37760	0	0	20	84	5	38	150	12	28	2	4	1	6	7	0	0	0	0	0	0	3	
20	6565	1949	Master	Married	76995	0	1	91	1012	80	498	0	16	176	2	11	4	9	5	0	0	0	1	0	0	3	
21	2278	1985	2n Cycle	Single	33812	1	0	86	4	17	19	30	24	39	2	2	1	3	6	0	0	0	0	0	0	3	
22	9360	1982	Graduation	Married	37040	0	0	41	86	2	73	69	38	48	1	4	2	5	8	0	0	0	0	0	0	3	
23	5376	1979	Graduation	Married	2447	1	0	42	1	1	1725	1	1	1	15	0	28	0	1	0	0	0	0	0	0	3	
24	1993	1949	PhD	Married	58607	0	1	63	867	0	86	0	0	19	3	2	3	9	8	0	1	0	0	0	0	3	
25	4047	1954	PhD	Married	65324	0	1	0	384	0	102	21	32	5	3	6	2	9	4	0	0	0	0	0	0	3	
26	1409	1951	Graduation	Married	40689	0	1	69	270	3	27	39	6	99	7	7	1	5	8	0	0	0	0	0	0	3	
27	7892	1969	Graduation	Single	18589	0	0	89	6	4	25	15	12	13	2	2	1	3	7	0	0	0	0	0	0	3	
28	2404	1976	Graduation	Married	53359	1	1	4	173	4	30	3	6	41	4	5	1	4	7	0	0	0	0	0	0	3	
29	5255	1986	Graduation	Single	51381.5	1	0	19	5	1	3	3	263	362	0	27	0	0	1	0	0	0	0	0	0	3	
30	9422	1989	Graduation	Married	38360	1	0	26	36	2	42	20	21	10	2	2	1	4	3	0	0	0	0	0	0	3	
31	1966	1965	PhD	Married	84618	0	0	96	684	100	801	21	66	0	1	6	9	10	2	0	0	1	0	0	0	3	
32	6864	1989	Master	Divorced	10979	0	0	34	8	4	10	2	2	4	2	3	0	3	5	0	0	0	0	0	0	3	
33	3033	1963	Master	Married	38620	0	0	56	112	17	44	34	22	89	1	2	5	3	3	0	0	0	0	0	0	3	
34	5710	1970	Graduation	Married	40548	0	1	31	110	0	5	2	0	3	2	2	1	4	5	0	1	0	0	0	0	3	
35	7373	1952	PhD	Divorced	46610	0	2	8	96	12	96	33	22	43	6	4	1	6	6	0	0	0	0	0	0	3	
36	8755	1946	Master	Married	68657	0	0	4	482	34	471	119	68	22	1	3	5	9	7	0	0	0	0	0	0	3	
37	10738	1951	Master	Single	49389	1	1	55	40	0	19	2	1	3	1	2	0	3	7	0	0	0	0	0	0	3	
38	4339	1970	PhD	Married	67353	0	1	37	702	17	151	0	8	35	5	5	6	12	2	0	0	0	0	0	0	3	
39	10755	1976	2n Cycle	Married	23718	1	0	76	6	3	14	15	7	36	3	3	1	2	7	0	0	0	0	0	0	3	
40	8595	1973	Graduation	Widow	42429	0	1	99	55	0	6	2	0	4	2	1	1	3	5	0	0	0	0	0	0	3	
41	2968	1943	PhD	Divorced	48948	0	0	53	437	8	206	160	49	42	2	7	10	5	6	1	0	0	0	0	0	3	

Our business questions

01

What are the major customer segments based on spending habits, de-
mographics and purchasing channels?

02

How does customer income, family composition, and engagement (Re-
cency, Web Visits, Complaints) impact campaign acceptance rates?

03

Which product categories contribute most to total spending within each customer segment, and how can promotions be tailored to increase cross-
category purchases?

Data Cleaning/handling

- Removed extra spaces in text columns.
- Converted numeric-looking strings to actual numbers.
- Replaced all missing numeric values (e.g., Income, Recency, etc.) with the same global median value across the entire dataset.
- Filled all missing categorical values with "Unknown".
- Repaired unrealistic Year_Birth entries (<1900 or >2005) by replacing them with the median year; derived Age = 2025 − Year_Birth.
- Created new aggregated features:

  TotalMnt = total spending across all product categories

  TotalPurchases = total purchases across all channels

  AcceptedCampaigns = total campaigns accepted
- Removed exact duplicate records

Variables

Dependent Variable (DV):

Response (1 = accepted campaign, 0 = not accepted)

Independent Variables (IVs)

Demographics: Age, Income, Kidhome, Teenhome

Behavior: Recency, Complain, AcceptedCampaigns

Spending: TotalMnt, MntWines, MntMeat, MntFruits, etc.

Engagement: NumDealsPurchases, NumWebPurchases, NumCatalogPurchases, NumStorePurchases, NumWebVisitsMonth

Excluded Variables

ID → identifier only

Z_CostContact, Z_Revenue → constant, no predictive value

Corelation

Purchasing channel totals are also highly correlated:
NumStorePurchases: TotalPurchases (0.82)
NumWebPurchases: TotalPurchases (0.78)
NumCatalogPurchases: TotalPurchases (0.74)

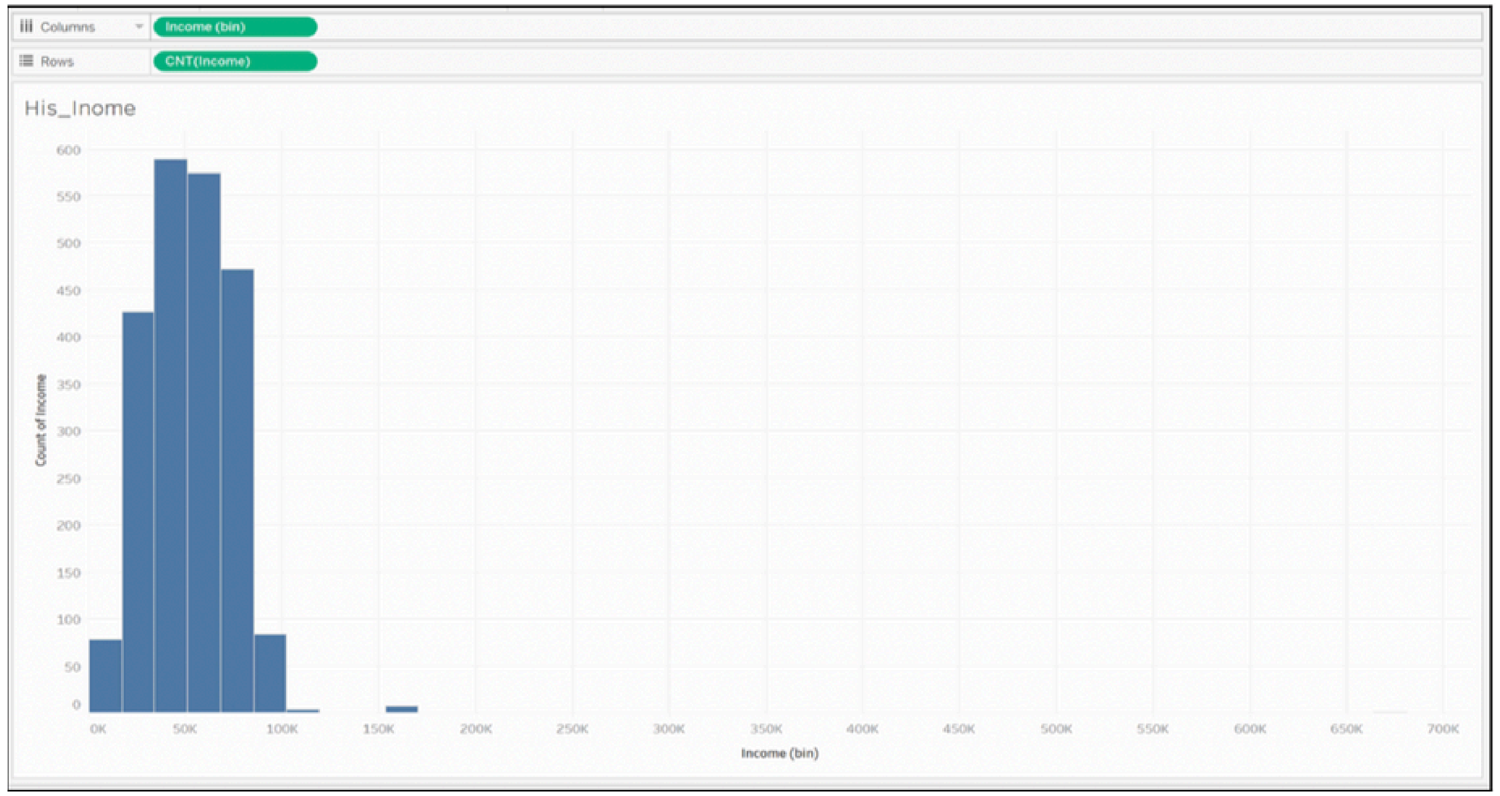
Strongest correlations are found between spending variables and their totals:
o MntWines: TotalMnt (0.89)
o MntMeatProducts: TotalMnt (0.84)

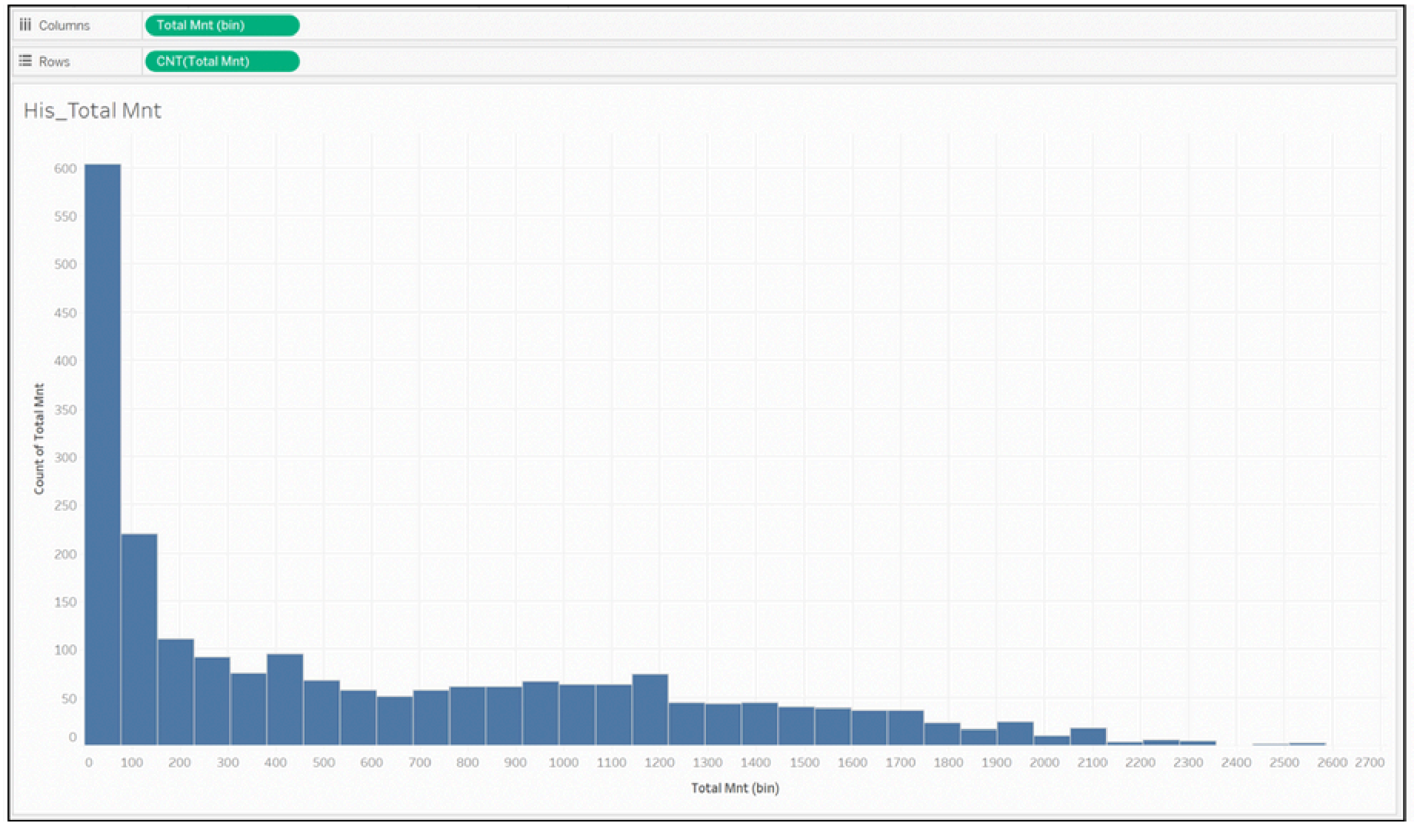
	ID	Year_Birth	Income	Kidhome	Teenhome	Recency	MntWines	MntFruits	tMeatProdu	ntFishProduc	tSweetProduc	MntGoldProds
ID	1											
Year_Birth	0.00121	1										
Income	0.013	-0.15934	1									
Kidhome	0.00241	0.229448	-0.4253	1								
Teenhome	-0.0026	-0.35793	0.019	-0.0361	1							
Recency	-0.0465	-0.02406	-0.0041	0.00883	0.016198	1						
MntWines	-0.0229	-0.15941	0.5769	-0.4963	0.004846	0.01606	1					
MntFruits	0.0046	-0.01219	0.4288	-0.3726	-0.17676	-0.0043	0.38964	1				
MntMeatPr	-0.0044	-0.02878	0.5778	-0.4371	-0.26116	0.02306	0.56267	0.543105	1			
MntFishPr	-0.0245	-0.04069	0.4376	-0.3876	-0.20419	0.00108	0.39975	0.594804	0.568402	1		
MntSweetP	-0.0076	-0.01795	0.4361	-0.3707	-0.16248	0.02267	0.38658	0.567164	0.523846	0.5798701	1	
MntGoldP	-0.0134	-0.05593	0.3219	-0.3496	-0.02173	0.01669	0.38752	0.392995	0.350609	0.4228748	0.3697243	1
NumDeals	-0.0372	-0.06572	-0.0823	0.2218	0.387741	-0.0011	0.01094	-0.13211	-0.12242	-0.1393606	-0.1201003	0.04908516
NumWebP	-0.0189	-0.15082	0.3806	-0.3616	0.1555	-0.0107	0.54226	0.296735	0.293761	0.2936807	0.3485443	0.4218359
NumCatalc	-0.0034	-0.12291	0.5868	-0.5022	-0.11077	0.02511	0.63523	0.487917	0.723827	0.5344783	0.4909239	0.43769744
NumStore	-0.0149	-0.13533	0.5266	-0.4997	0.050695	0.0008	0.6421	0.461758	0.479659	0.4598546	0.448756	0.38167801
NumWebV	-0.0074	0.116822	-0.5498	0.44785	0.134884	-0.0214	-0.32065	-0.41838	-0.53947	-0.4460027	-0.4232941	-0.25071874
AcceptedC	-0.036	0.06125	-0.0161	0.01467	-0.04268	-0.033	0.0622	0.014727	0.018272	0.0003572	0.0015299	0.12309106
AcceptedC	-0.0254	-0.06313	0.1827	-0.1616	0.038886	0.01883	0.37329	0.010152	0.102912	0.0168428	0.0286407	0.02226798
AcceptedC	-0.0075	0.016029	0.3349	-0.2056	-0.19105	0.00013	0.47261	0.215833	0.373769	0.1995778	0.2595904	0.18102116
AcceptedC	-0.0216	-0.0075	0.2749	-0.1723	-0.14009	-0.0193	0.35413	0.194748	0.309761	0.2607621	0.2418178	0.1663961
AcceptedC	-0.0151	-0.0073	0.0876	-0.0817	-0.0156	-0.0018	0.20591	-0.00977	0.043033	0.0025767	0.0099852	0.04999029
Complain	0.03388	-0.03118	-0.0272	0.04021	0.003138	0.01323	-0.03901	-0.00517	-0.02348	-0.0209528	-0.022485	-0.03086123
Response	-0.022	0.019339	0.1329	-0.08	-0.15445	-0.1984	0.24725	0.125289	0.236335	0.1113308	0.1173719	0.13985014
Age	-0.0012	-1	0.1593	-0.2294	0.357933	0.02406	0.15941	0.012188	0.02878	0.040687	0.0179453	0.05593276
TotalMnt	-0.0181	-0.11045	0.6648	-0.5567	-0.13838	0.02043	0.89184	0.614229	0.842965	0.6428182	0.6030164	0.52426185
TotalPurch	-0.0238	-0.17524	0.5635	-0.4779	0.133163	0.00574	0.71279	0.455461	0.554229	0.469454	0.4728759	0.49331423
AcceptedC	-0.0369	0.001199	0.3071	-0.2119	-0.12776	-0.0126	0.51236	0.161002	0.309313	0.1778296	0.2002372	0.19374081
Education	-0.0004	0.094639	-0.0574	0.01973	-0.05647	-0.0075	-0.0992	0.021065	-0.03594	0.0575531	0.0549832	0.01438814
Education	-0.0095	0.114607	-0.2005	0.05414	-0.11942	-0.0036	-0.13857	-0.06004	-0.10831	-0.0589062	-0.0569401	-0.06385138
Education	0.01871	0.065205	0.0189	0.00065	-0.02221	0.03218	-0.05881	0.113177	0.055913	0.103627	0.1049492	0.13172241
Education	-0.0258	-0.07526	0.0117	0.00815	0.023601	-0.0234	0.03852	-0.052	-0.00704	-0.0441865	-0.0634521	-0.03091256
Education	0.00437	-0.11985	0.0811	-0.042	0.089451	-0.0114	0.1573	-0.08277	0.003855	-0.1040597	-0.0872484	-0.11819559
Marital_St	-0.0173	-0.0749	0.0081	-0.0192	0.052613	0.00443	0.02112	0.009612	-0.02522	-0.0154489	-0.0020094	0.01477391
Marital_St	0.0192	0.003348	0.0062	0.02201	0.029078	-0.0043	-0.00574	-0.02173	-0.01843	-0.0113939	-0.0192441	-0.02347009
Marital_St	-0.0192	0.124579	-0.0271	0.02127	-0.09264	0.00157	-0.02516	0.003882	0.031798	0.0034761	-0.000329	-0.00406516
Marital_St	0.02194	-0.16516	0.0315	-0.0738	0.045093	0.00022	0.03663	0.032211	0.018675	0.0478944	0.0546309	0.04610336

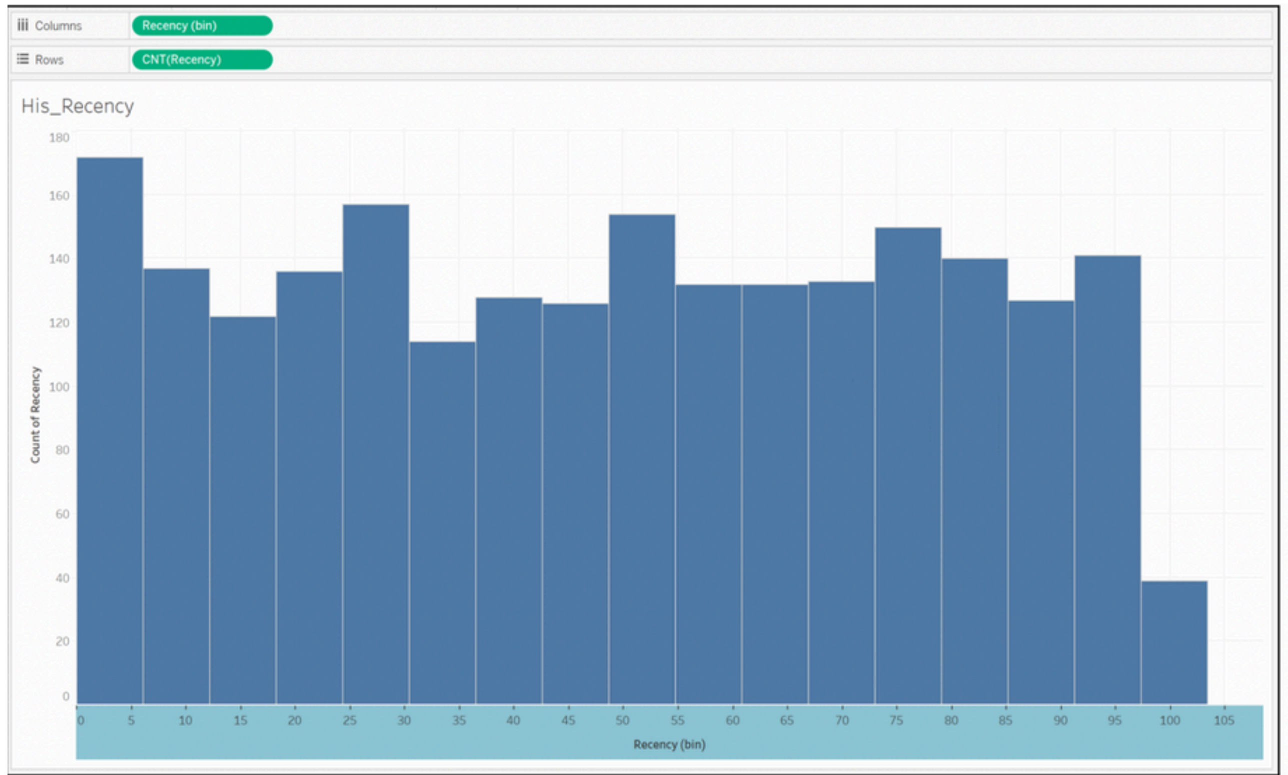
	umDealsPurchas	umWebPurchas	CatalogPurch	StorePurch	WebVisitsM	AcceptedCmp3	AcceptedCmp4	AcceptedCmp5	AcceptedCmp1	AcceptedCmp2	Complain	Response	Age
NumDeals	1												
NumWebP	0.234184688	1											
NumCatalc	-0.008617246	0.37837555	1										
NumStore	0.068878826	0.50271341	0.5187383	1									
NumWebV	0.347633358	-0.0558463	-0.5203637	-0.42847	1								
AcceptedC	-0.023108937	0.04217553	0.1047295	-0.06758	0.06121	1							
AcceptedC	0.015594262	0.15590282	0.1392473	0.17938	-0.0321	-0.07951227	1						
AcceptedC	-0.183247862	0.13868404	0.3223215	0.21479	-0.2781	0.080315817	0.306525764	1					
AcceptedC	-0.123244434	0.15514283	0.3080966	0.18325	-0.1925	0.094750533	0.251299934	0.403077676	1				
AcceptedC	-0.037694758	0.03418774	0.0998524	0.08519	-0.0072	0.072019919	0.292209863	0.221533005	0.17531458	1			
Complain	0.000419649	-0.0163102	-0.0204531	-0.01652	0.01977	0.008415398	-0.02761147	-0.009418558	-0.02549862	-0.01133432	1		
Response	0.002238313	0.14872959	0.2208104	0.03936	-0.004	0.254258283	0.177018602	0.326633945	0.29398153	0.16929266	-0.00171	1	
Age	0.065722938	0.15081964	0.1229144	0.13533	-0.1168	-0.06125019	0.063126581	-0.016029266	0.00750071	0.00730347	0.03118	-0.0193	1
TotalMnt	-0.065111944	0.51983657	0.7785769	0.67467	-0.5002	0.053384768	0.253290177	0.470058408	0.38152284	0.13581286	-0.03706	0.2653	0.11045
TotalPurch	0.362334293	0.77783124	0.7351983	0.82026	-0.3123	0.020708859	0.189331426	0.217754249	0.22004081	0.07697997	-0.02058	0.15514	0.17524
AcceptedC	-0.123949484	0.19154365	0.3458223	0.20656	-0.1663	0.42942609	0.614631595	0.715771239	0.67932166	0.45847962	-0.02222	0.42604	-0.0012
Education	-0.012859787	-0.0404259	-0.0369275	-0.02747	0.01779	0.001365731	-0.03039724	-0.028568109	0.00602353	-0.00972357	0.03384	-0.0361	-0.0946
Education	-0.043015068	-0.124235	-0.1172723	-0.14209	0.10066	0.023200467	-0.04460981	-0.044029865	-0.04119622	-0.01831202	-0.01529	-0.0495	-0.1146
Education	-0.007984449	0.01459349	0.0219566	0.01222	-0.012	-0.01378171	-0.01027205	0.013717259	0.03476682	0.0070393	0.03182	-0.0402	-0.0652
Education	0.023492636	-0.0083883	-0.01397	0.01466	-0.0253	-0.01353342	0.015630824	0.004979472	-0.02836001	-0.03090862	-0.01832	0.00618	0.07526
Education	0.013483996	0.06424158	0.055311	0.04398	-0.0124	0.019325736	0.036149845	0.015156053	-0.00548857	0.03289403	-0.03997	0.08676	0.11985
Marital_St	0.019415589	0.0275935	0.001205	0.00301	0.0239	0.01758597	0.003924393	-0.02189685	-0.01740914	0.02412588	-0.00266	0.05515	0.0749
Marital_St	0.027561174	-0.000219	-0.004576	0.00643	0.00051	-0.01876312	-0.00641127	0.020758451	0.01157866	-0.0029721	-0.00539	-0.1483	-0.0033
Marital_St	-0.046920557	-0.036339	-0.0145003	-0.02576	-0.0034	0.015470992	-0.01303344	-0.017911329	-0.00078918	-0.01410193	0.01634	0.10858	-0.1246
Marital_St	0.001236709	0.03657635	0.0427795	0.0363	-0.0337	-0.01512271	0.039734009	0.022610373	0.00049954	-0.00066614	-0.01835	0.05172	0.16516

	TotalMnt	otalPurchase	eptedCampaig	Education_2n Cycle	ucation_Ba	ucation_Graduati	ducation_Maste	Education_PhD	_Status_Div	l_Status_Ma	Status_	Status_Widow
TotalMnt	1											
TotalPurch	0.7539	1										
AcceptedC	0.45955	0.257273	1									
Education	-0.0573	-0.04356	-0.0216605	1								
Education	-0.1368	-0.160612	-0.0432553	-0.049616292	1							
Education	0.02357	0.016805	0.0097625	-0.317663181	-0.1582	1						
Education	0.00442	0.003765	-0.0127166	-0.140421162	-0.0699	-0.44760449	1					
Education	0.05823	0.066328	0.0307968	-0.166171234	-0.0827	-0.529685051	-0.234144197	1				
Marital_St	0.00273	0.016608	-0.0023377	0.010079657	-0.0439	0.006666441	-0.005213678	0.00591585	1			
Marital_St	-0.0159	0.007839	0.0019645	0.02261441	-0.0052	-0.014034267	0.015460842	-0.01071062	-0.4587	1		
Marital_St	-0.0019	-0.04139	-0.0086578	-0.026252767	0.04458	0.019476769	-0.012001307	-0.01111751	-0.1787	-0.70942	1	
Marital_St	0.04168	0.045212	0.0183239	-0.01688436	-0.0137	-0.018331965	-0.004742754	0.043360306	-0.0641	-0.25462	-0.099	1

Histograms

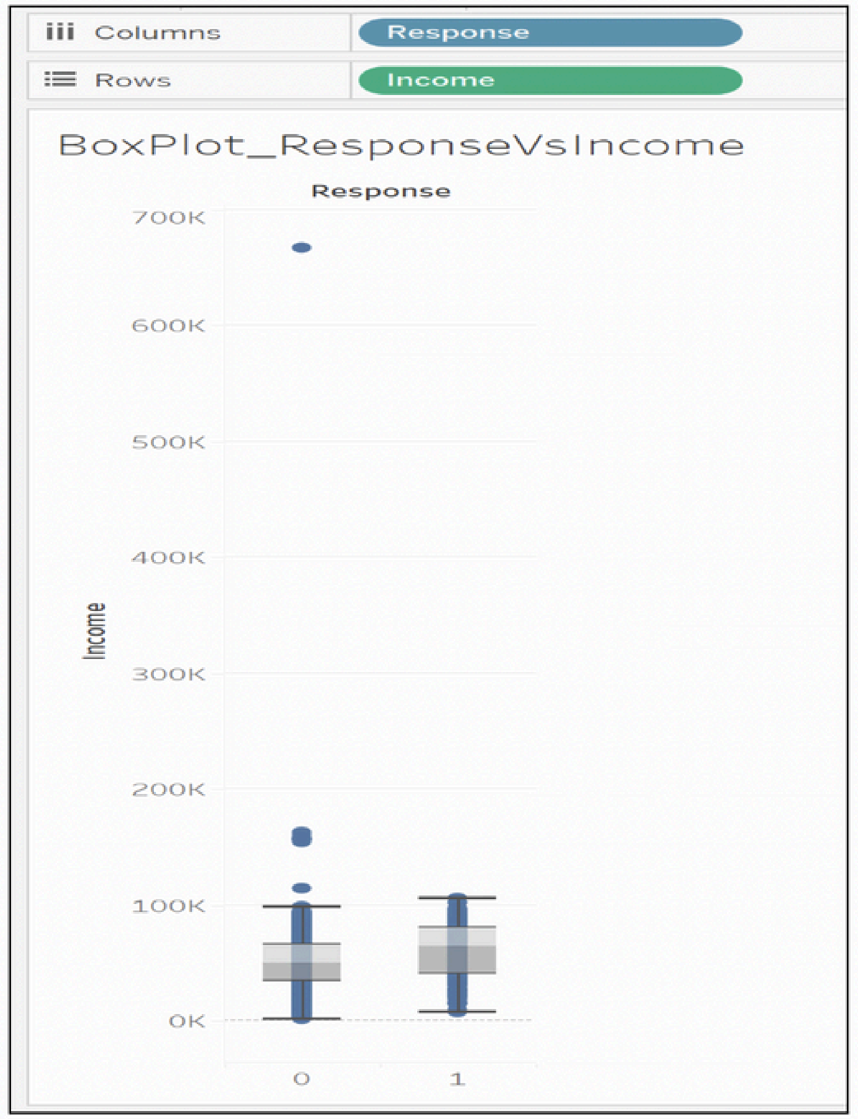

Columns
Income (bin)
Rows
CNT(Income)
His_Inome
Count of Income
Income (bin)
0K
50K
100K
150K
200K
250K
300K
350K
400K
450K
500K
550K
600K
650K
700K


Columns
Total Mnt (bin)
Rows
CNT(Total Mnt)
His_Total Mnt
Count of Total Mnt
Total Mnt (bin)

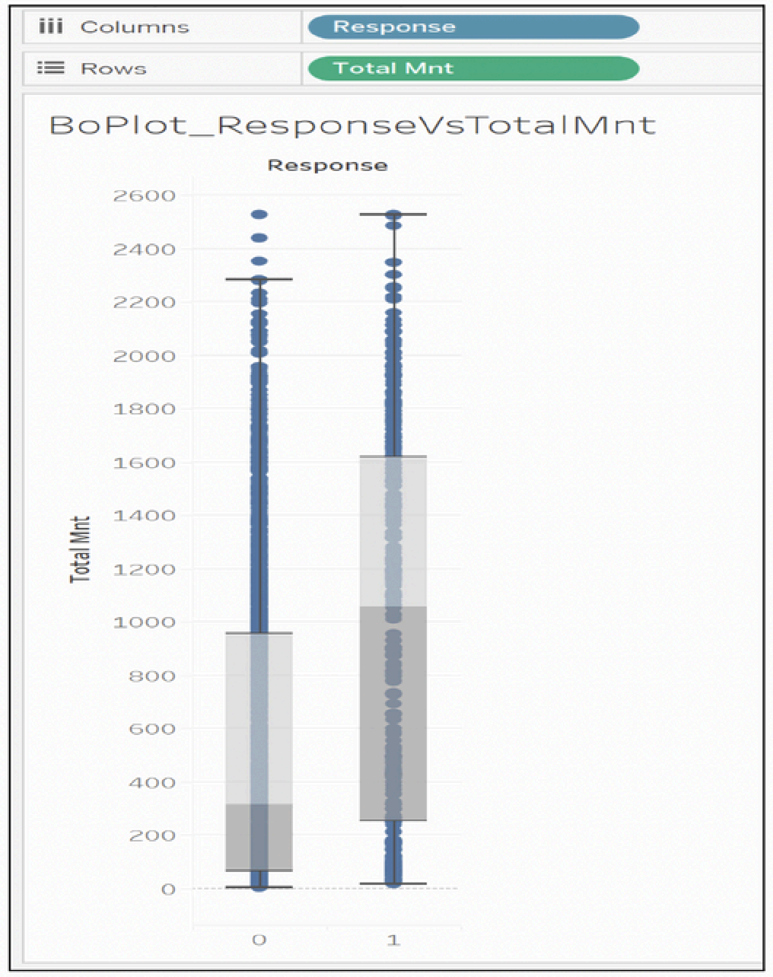

Columns
Recency (bin)
Rows
CNT(Recency)
His_Recency
Count of Recency
Recency (bin)

Boxplots

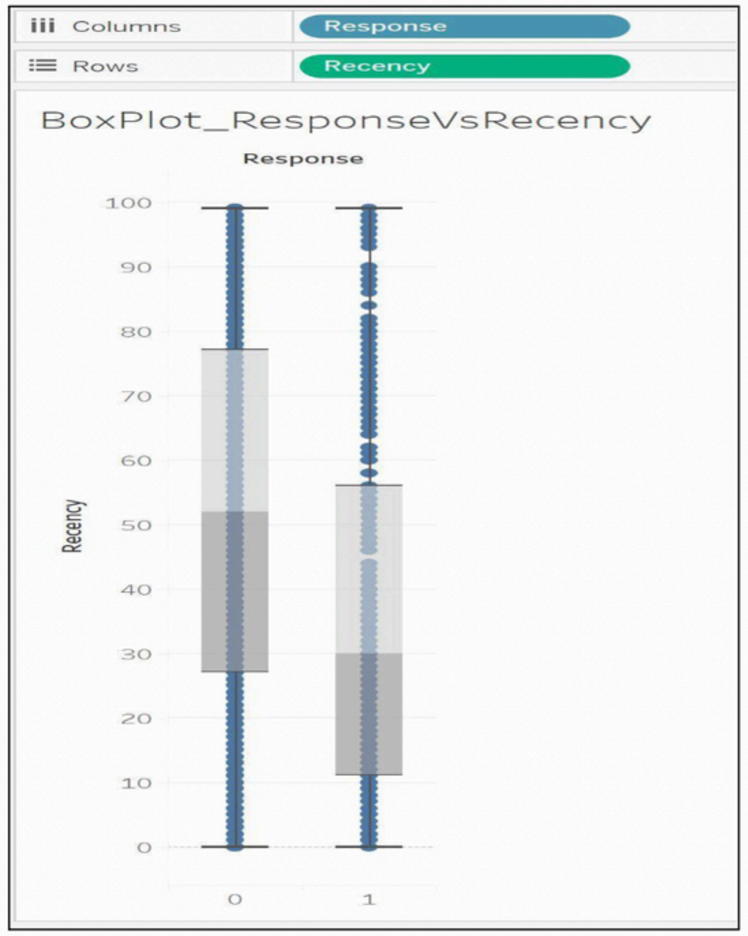
1. BoxPlot: Response vs Income



2. BoxPlot: Response vs Total Mnt (Total Spending)



3. BoxPlot: Response vs Recency



Logistic Regression

Purpose: Predict which customers will accept a marketing campaign (Response = 1).

Why Logistic Regression?

- *Target variable is binary*
- *Highly interpretable for managers*
- *Works well with mixed variables (numeric + categorical)*
- *Provides probability scores for targeting decisions*

How We Ran the Model?

Process

- *Cleaned data and created key features (TotalMnt, TotalPurchases, AcceptedCampaigns, Age)*
- *Converted categorical variables into dummy variables*
- *Used stepwise logistic regression to select significant predictors*
- *Split data into training & validation sets*
- *Evaluated performance using ROC curves, decile charts, and classification metrics*

Final Predictors:

- Teenhome, Recency,NumCatalogPurchases, NumWebVisitsMonth,AcceptedCmp3, AcceptedCampaigns, TotalMnt,Education_PhD, Marital_Status_Married

Performance

Training AUC: 0.87

Validation AUC: 0.869

→ Strong, stable model with no overfitting.

Default Cutoff (0.50)

Sensitivity: 0.36

Precision: 0.64

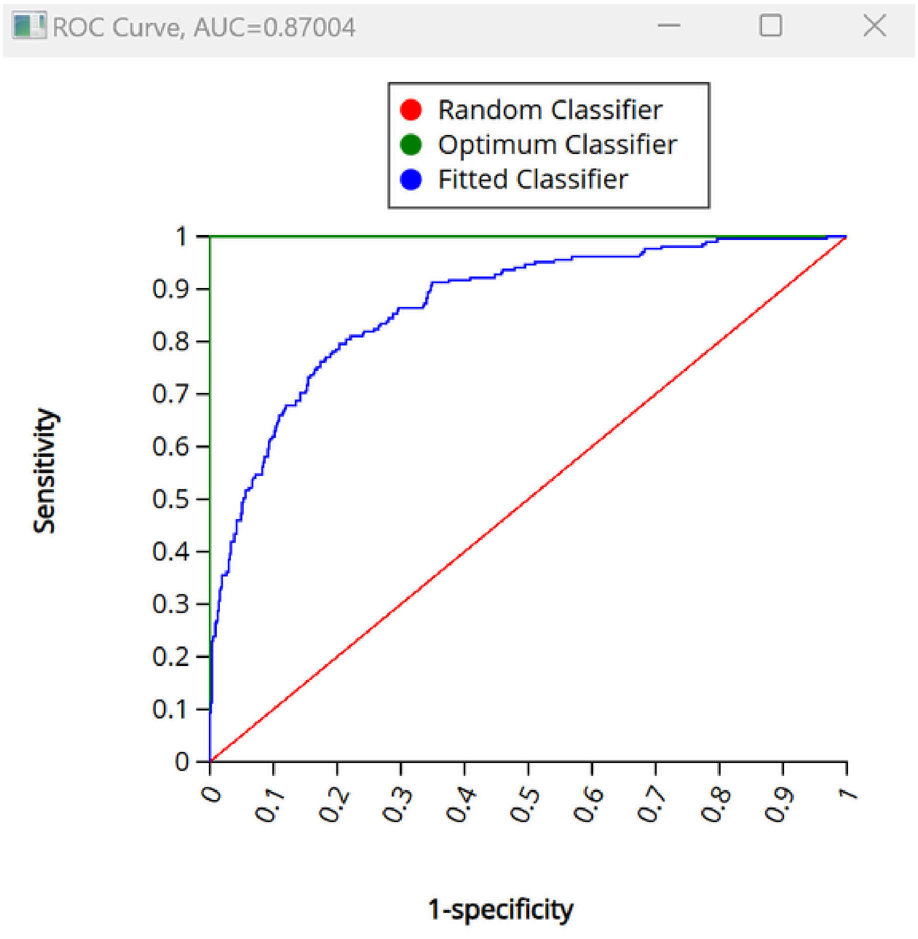
Optimal Cutoff (0.22)

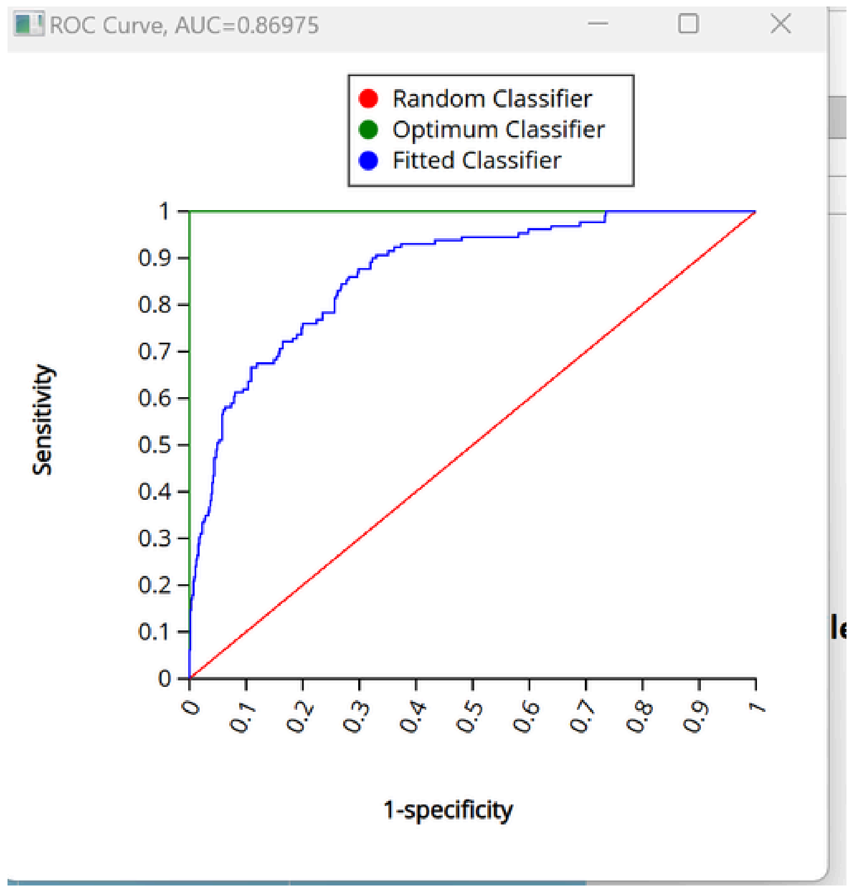
Sensitivity: 0.67

Precision: 0.48

Highest F1 Score: 0.56

→ Best balance for marketing objectives (reach more responders).





Key Predictors:

Strong positive: AcceptedCampaigns, AcceptedCmp3, Web visits, Catalog purchases, TotalMnt

Negative: Recency, Teenhome, Married status

How does it answer our business questions?

Q2: Can we predict who will respond?

Yes.

AUC ≈ 0.87 → strong predictive power

Model assigns a probability score to every customer

Allows ranking of customers from most likely to least likely to respond

Q1: Which customer characteristics are associated with response?

<u>More likely to respond:</u>

Previously accepted campaigns

High website activity

Catalog shoppers

High spenders

PhD-educated customers

<u>Less likely to respond:</u>

Married customers

Teenhome households

Customers with high recency (inactive)

Q3: How do we select who to target?

Using the cutoff = 0.22:

Captures most responders (higher sensitivity)

Creates a high-probability target list

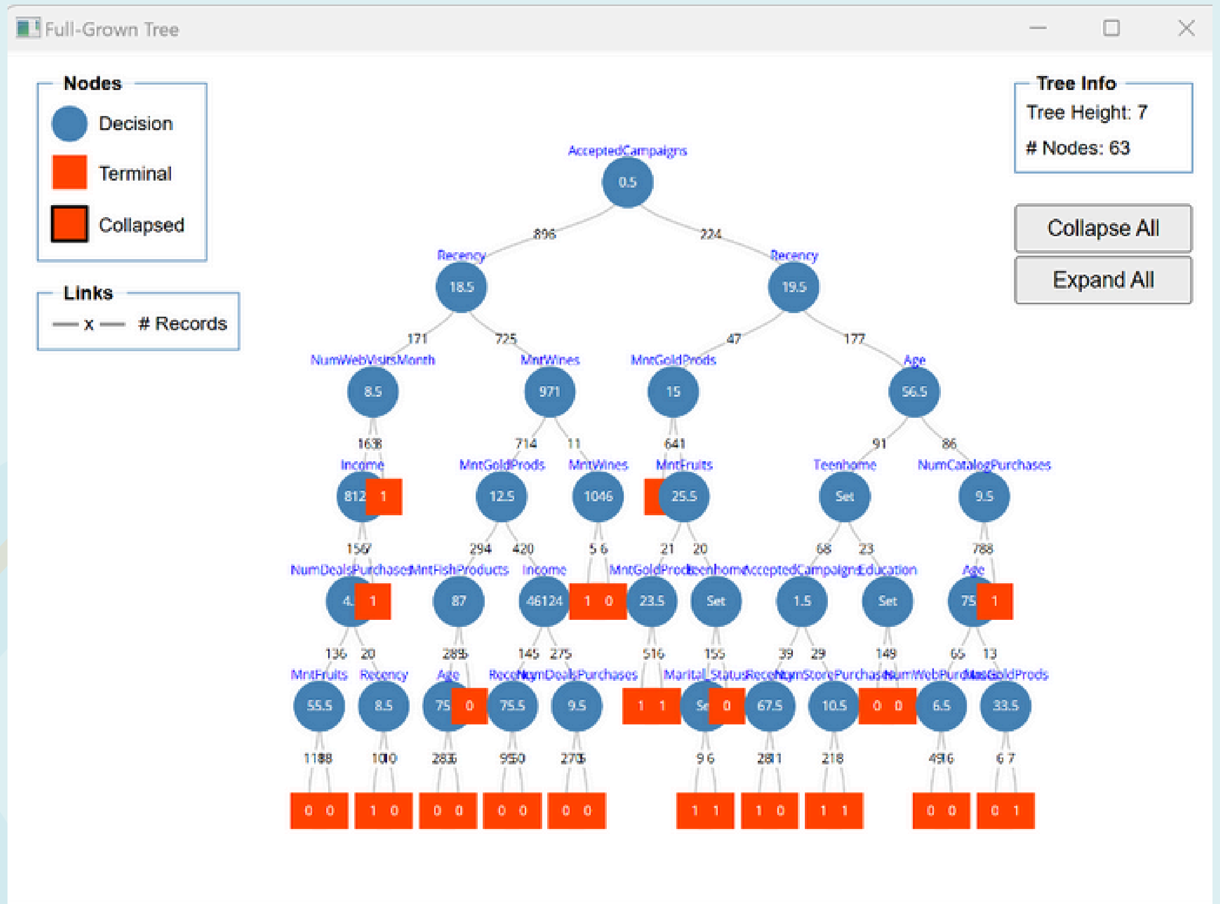
Focuses marketing efforts on customers with stronger engagement and past acceptance

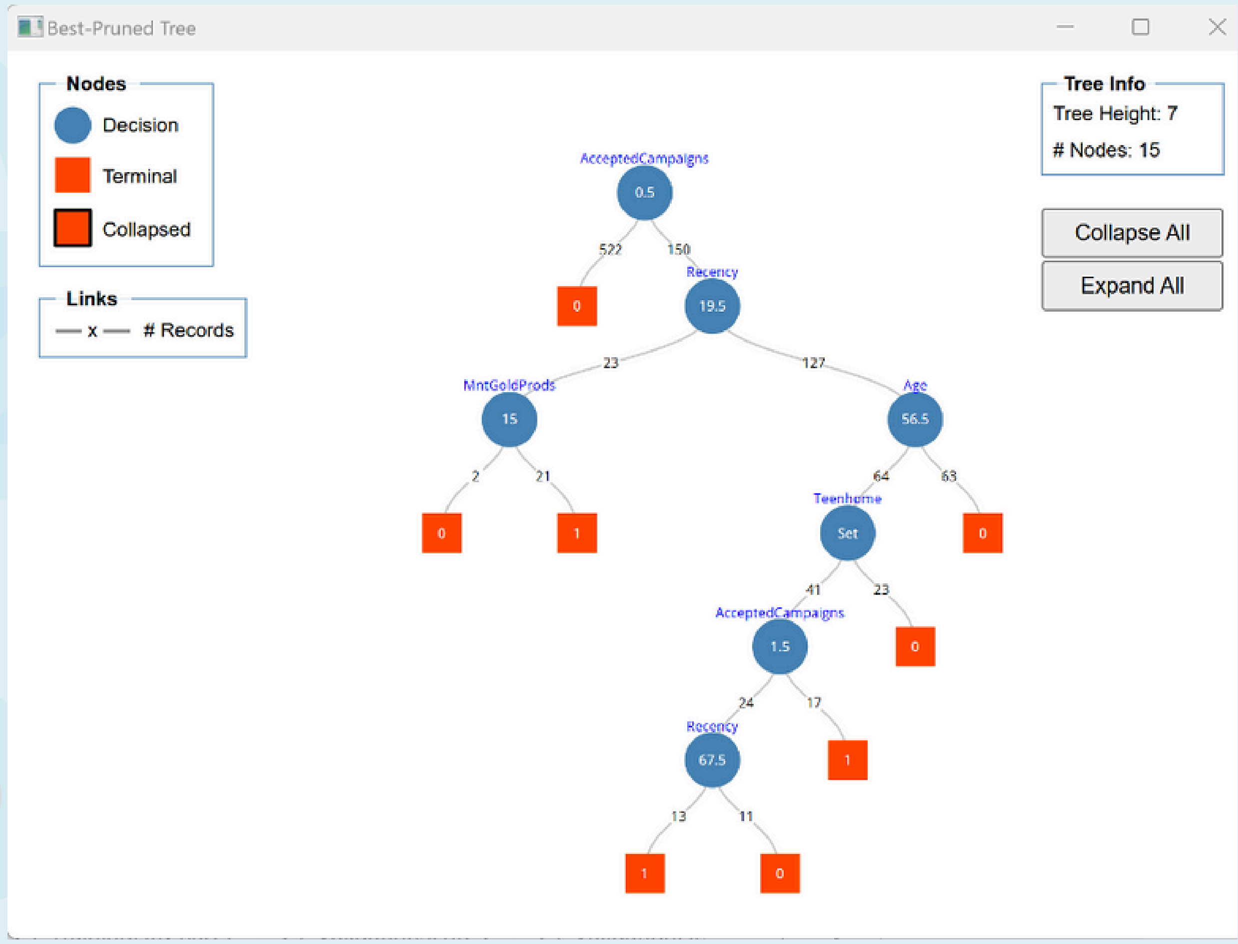
Decision Tree Model Overview

Models Built:
Full-Grown Tree
Best-Pruned Tree

Why Decision Trees?

- Handle numeric + categorical variables
- Highly interpretable (clear business rules)
- Show natural segmentation of customers
- Useful for binary classification (Response 0/1)

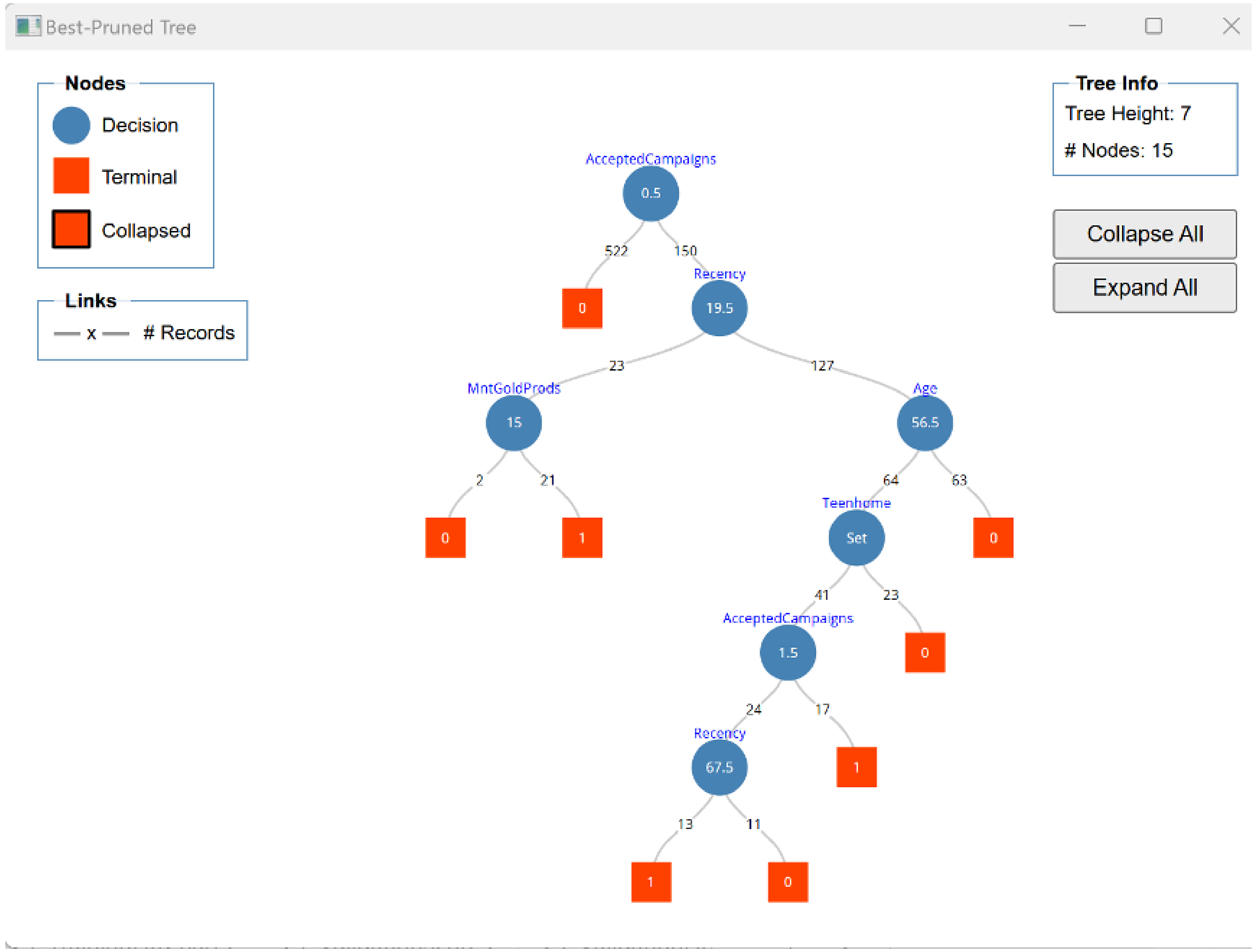




How We Ran the Model?
Process

- Removed irrelevant variables (ID, constants)
- Used multiple terminal-node settings (5, 10, 20, 30)
- Trained full tree → pruned for lowest validation error
- Evaluated using ROC/AUC on validation & test sets
- Selected optimal cutoff using top 20% probabilities (cutoff = 0.17)
- Final Model Used:
- → Best-Pruned Tree (5-terminal-node setting)

If AcceptedCampaigns ≤ 0.5 → No Response
If AcceptedCampaigns > 0.5 AND Recency ≤ 19.5 → Response
If Age > 56.5 → No Response
If MntGoldProds > 2.5 → Response
If Teenhome = 1 AND AcceptedCampaigns > 1.5 AND Recency ≤ 51.5 → Response
Otherwise → No Response



Model Performance

Performance Metrics

Validation AUC ≈ 0.74

Test AUC ≈ 0.68

→ Stable performance, no major overfitting

Cutoff Choice

Default (0.50):

• Very high specificity / low sensitivity (misses responders)

Optimal (0.17):

• Sensitivity doubles (0.28 → 0.54)

• F1 Score improves (0.39 → 0.54)

→ Better for marketing where capturing responders matters most

How does it answer our business questions?

What customer segments exist?

Tree reveals clear segments:

Low Recency = high response

Heavy website visitors + low purchases = low response

Households with teens + low spending = low response

High-spending customers = moderate response

Do income, family composition, or engagement matter?

Engagement (Recency, Web Visits) = strongest predictors

Family composition (Teenhome) reduces response

Income does not matter → excluded automatically by the tree

Complaints associated with low response

3. Cross-Selling Opportunities?

Higher spenders (TotalMnt, especially Wines/Gold products) appear in higher-response nodes → good targets for premium product promotion

Low-spending families better for value offers

Neural Network Model Overview

- Built multiple neural network architectures
- Goal: Predict campaign response (0/1)
- Final chosen model: Net 11 (1 hidden layer, 11 nodes)
- Neural networks capture complex non-linear patterns that other models mis

How We Ran the Neural Network?

- One-hot encoded all categorical variables
- Standardized all numerical inputs (z-score)
- Removed meaningless variables (ID, constants, Year_Birth)
- Trained several architectures (1-layer, 2-layer, various node sizes)
- Compared AUC, accuracy, lift, and stability
- Selected model based on best validation performance

Model	Hidden Layers	Nodes	Iterations	AUC-ROC	Status
Net 11	**1 layer**	**11**	**243**	**809**	**BEST**
Net 28	2 layers	1, 7	351	800	Good
Net 16	1 layer	16	199	786	Good
Net 10	1 layer	10	245	775	Moderate
Net 13	1 layer	13	205	772	Moderate
Net 17	1 layer	17	221	722	Overfitting
Net 92-96	2 layers	5, 11-15	205	#N/A	Failed

Why Net 11?

Highest validation AUC: 0.809

Avoided overfitting (unlike deeper models)

Best balance between complexity and performance

Model Performance

AUC-ROC
Train: 0.85
Validation: 0.809
Test: 0.809
Strong and stable prediction across datasets
Accuracy
Training: 0.82
Validation: 0.79
Test: 0.78
Lift Chart Highlights
Top 10% customers → 3.8× lift
Top 20% customers → 64% of total responders captured
Neural network outperforms the Decision Tree (AUC ≈ 0.68).

Tested cutoffs: 0.185, 0.20, 0.229
Recommended Cutoff: 0.20
Best F1-score: 0.486
Sensitivity: 60.5%
Specificity: 85.1%
Accuracy: 81.58%
Why 0.20?
Captures more true responders
Balanced precision + recall
Ideal for marketing where missing responders is costly

How does it answer our business questions?

Customer Segments Identified:
High Response: Past acceptors, recent shoppers, high spenders
Medium Response: Catalog shoppers, premium product buyers
Low Response: Inactive, low spenders, high web-visits but no purchases

Impact of Demographics & Behavior:
Engagement (Recency, AcceptedCampaigns) → strongest impact
Families with kids/teens → lower response
Income → minimal predictive power
Complaints → reduces response likelihood

Cross-Selling Opportunities:
Premium-category buyers (Gold, Wines, Meat) → ideal for upselling
Catalog shoppers → best channel for promotion
Multi-category spenders → suited for bundles

So, which is the best model?

- Compared three models: Logistic Regression, Decision Tree, Neural Network
- Logistic Regression achieved the highest ROC/AUC (~0.87)
- Neural Network performed well (AUC ~0.809)
- Decision Tree performed moderately (AUC ~0.80)
- ROC/AUC chosen as main metric because it evaluates performance across all cutoffs
- Final Selected Model: Logistic Regression
- Most accurate
- Most stable
- Easiest to interpret and explain for business use

Recommendations

- Target customers with predicted probability ≥ 0.25
- Prioritize:

  High spenders

  Recently active customers

  Customers who accepted previous campaigns

  Strengthen web and catalog buying channels

Project Learnings

Data preprocessing choices strongly influence model performance

Choosing the right cutoff is essential for marketing accuracy

Simple, interpretable models often outperform black-box methods

ROC curves + lift charts are powerful for model comparison

Thank you