Superstore Marketing Campaign Analysis – From Raw Data to Insights

Customers: 2240; Avg Income: 52247.25; Avg Spend: 605.8

When I first started this project, I knew I wanted to analyze how marketing campaigns influence customer behavior and sales in a retail environment. The dataset was inspired by a fictional Superstore Marketing Campaign scenario, bringing together sales, customers, products, and campaign performance.
But before I could even think about running analytics or building dashboards, there was one unavoidable step: Data Cleaning.

Data Cleaning

The raw dataset wasn't ready to use. It had:

- Missing customer demographic values (age, gender, segment)
- Duplicate order entries
- Inconsistent formats for dates and regions
- Columns with unnecessary spaces in names

I started with the following actions:

- Removed duplicate rows
- Standardized column names (e.g., Order Date → order_date)
- Filled missing values using logical imputation (e.g., median for Age, mode for Segment)
- Converted all dates into proper datetime format
- Created derived columns like TotalAmount = Quantity * UnitPrice

Merging Data

Next, I brought together:

- Sales Data → Orders, quantities, revenue
- Customer Data → Age, gender, segment, loyalty score
- Campaign Data → Campaign IDs, budgets, duration, channel
- Engagement Data → Click rates, open rates, conversions

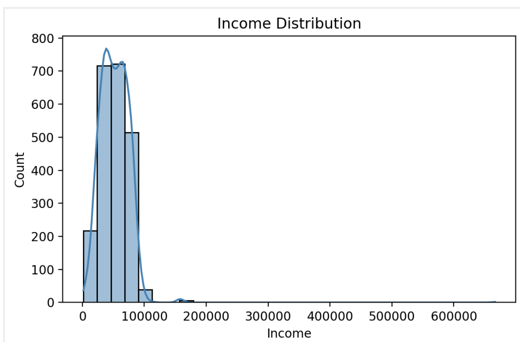
Prepared Dataset

After cleaning and merging, the dataset looked something like this:.

	Id	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	MntFruits	MntMeatProduc
2	1826	1970	Graduation	Divorced	84835	0	0	2014-06-16	0	189	104	379
3	1	1961	Graduation	Single	57091	0	0	2014-06-15	0	464	5	64
4	10476	1958	Graduation	Married	67267	0	1	2014-05-13	0	134	11	59
5	1386	1967	Graduation	Together	32474	1	1	2014-11-05	0	10	0	1
6	5371	1989	Graduation	Single	21474	1	0	2014-08-04	0	6	16	24
7	7348	1958	PhD	Single	71691	0	0	2014-03-17	0	336	130	411
8	4073	1954	2n Cycle	Married	63564	0	0	2014-01-29	0	769	80	252
9	1991	1967	Graduation	Together	44931	0	1	2014-01-18	0	78	0	11
10	4047	1954	PhD	Married	65324	0	1	2014-11-01	0	384	0	102
11	9477	1954	PhD	Married	65324	0	1	2014-11-01	0	384	0	102
12	2079	1947	2n Cycle	Married	81044	0	0	2013-12-27	0	450	26	535
13	5642	1979	Master	Together	62499	1	0	2013-09-12	0	140	4	61
14	10530	1959	PhD	Widow	67786	0	0	2013-07-12	0	431	82	441
15	2964	1981	Graduation	Married	26872	0	0	2013-10-16	0	3	10	8
16	10311	1969	Graduation	Married	4428	0	1	2013-05-10	0	16	4	12
17	837	1977	Graduation	Married	54809	1	1	2013-11-09	0	63	6	57
18	10521	1977	Graduation	Married	54809	1	1	2013-11-09	0	63	6	57
19	10175	1958	PhD	Divorced	32173	0	1	2013-01-08	0	18	0	2
20	1473	1960	2n Cycle	Single	47823	0	1	2013-07-23	0	53	1	5
21	2795	1958	Master	Single	30523	2	1	2013-01-07	0	5	0	3
22	2285	1954	Master	Together	36634	0	1	2013-05-28	0	213	9	76

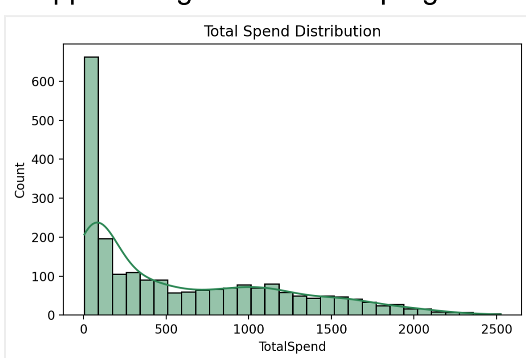
Who are our customers?

Superstore serves a diverse base across education and marital statuses, with varying income levels. Higher income correlates with higher spend, but the relationship is not perfectly linear, leaving room for marketing influence.



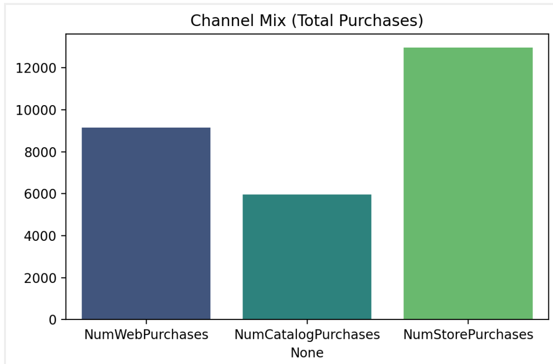
What do they buy and how much?

Spend is right-skewed. A minority of customers drive a disproportionate share of revenue, a classic long-tail pattern. This supports targeted retention programs for top-value customers.



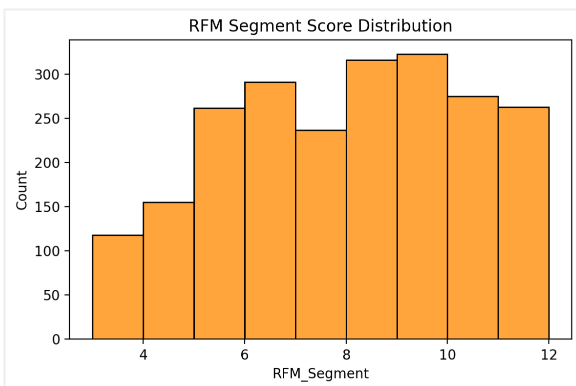
Where do purchases happen?

Store, Web, and Catalog each contribute, with opportunities to shift lower-cost channels for suitable segments. Understanding channel preference helps tailor offers and timing.

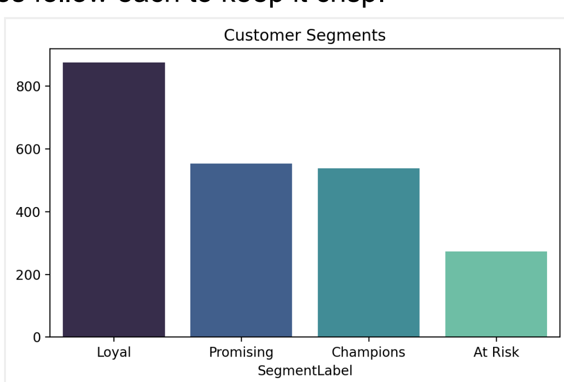


Segmenting with RFM

We segmented customers using Recency, Frequency, and Monetary value. Champions purchase recently, frequently, and spend the most; At Risk customers have lapsed or buy infrequently.

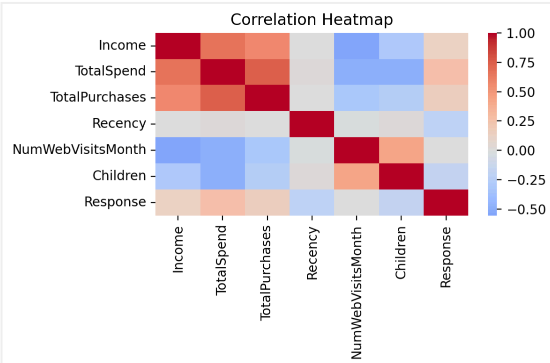


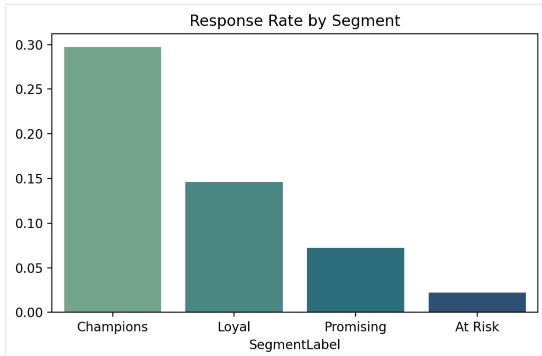
Here are the main results and visuals from the analysis I ran earlier. I'm showing the key tables and figures first; short notes follow each to keep it crisp.



What drives marketing response?

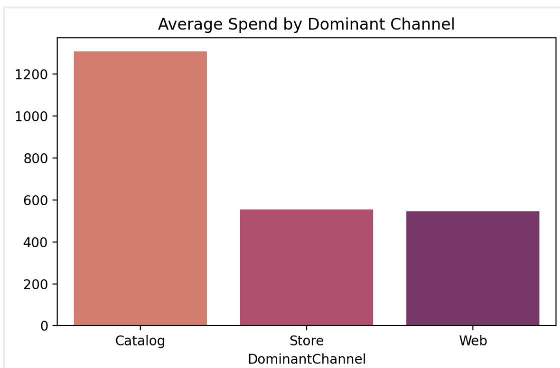
We explored correlations between response and key drivers like spend, frequency, site visits, and recency. While correlations are modest, segment-level differences are clear: higher-value segments respond more.





Channel strategy recommendations

Customers whose dominant channel is Web or Store show differing average spend.



Action Playbook

- Champions: early access and loyalty tiers; protect experience.
- Loyal: cross-sell bundles and increase frequency with personalized offers.
- Promising: onboarding sequences, welcome discounts, reduce friction to second purchase.
- At Risk: win-back campaigns based on last category purchased and preferred channel.

Data columns summarized:

Income, TotalSpend, TotalPurchases, Recency, NumWebVisitsMonth, and more from the source file.