Marketing Campaign Analytics & Customer Segmentation

**Executive Summary**

This project delivers a comprehensive marketing analytics solution for a grocery retailer, leveraging RFM segmentation, segment profiling, campaign ROI analysis, and predictive modelling to optimise campaign effectiveness and customer engagement.

**1. Business Background & Objectives**

* **Business context:** Rising competition and the need for targeted marketing in grocery retail
* **Objectives:**
  + Segment customers to maximise revenue and retention
  + Identify most responsive and valuable segments for campaigns
  + Predict future campaign responders for efficient targeting

**2. Data Overview**

* **Source:** [Kaggle Marketing Campaign Data](https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis)
* **Size:** ~2,200 customers, 29 features
* **Features include:**
  + Demographics (Year\_Birth, Education, Marital\_Status, Income, Country)
  + Family/household (Kidhome, Teenhome)
  + Date customer joined (Dt\_Customer)
  + Product spend (MntWines, MntMeatProducts, etc.)
  + Channel usage (NumWebPurchases, NumCatalogPurchases, NumStorePurchases)
  + Campaign response (AcceptedCmp1–5, Response)

**Data Dictionary**

|  |  |
| --- | --- |
| Column | Description |
| ID | Customer unique ID |
| AcceptedCmp1 | 1 if customer accepted the offer in the 1st campaign, 0 otherwise |
| AcceptedCmp2 | 1 if customer accepted the offer in the 2nd campaign, 0 otherwise |
| AcceptedCmp3 | 1 if customer accepted the offer in the 3rd campaign, 0 otherwise |
| AcceptedCmp4 | 1 if customer accepted the offer in the 4th campaign, 0 otherwise |
| AcceptedCmp5 | 1 if customer accepted the offer in the 5th campaign, 0 otherwise |
| Response (target) | 1 if customer accepted the offer in the last campaign, 0 otherwise |
| Complain | 1 if customer complained in the last 2 years |
| DtCustomer | date of customer’s enrolment with the company |
| Education | customer’s level of education |
| Marital | customer’s marital status |
| Kidhome | number of small children in customer’s household |
| Teenhome | number of teenagers in customer’s household |
| Income | customer’s yearly household income |
| MntFishProducts | amount spent on fish products in the last 2 years |
| MntMeatProducts | amount spent on meat products in the last 2 years |
| MntFruits | amount spent on fruits products in the last 2 years |
| MntSweetProducts | amount spent on sweet products in the last 2 years |
| MntWines | amount spent on wine products in the last 2 years |
| MntGoldProds | amount spent on gold products in the last 2 years |
| NumDealsPurchases | number of purchases made with discount |
| NumCatalogPurchases | number of purchases made using catalogue |
| NumStorePurchases | number of purchases made directly in stores |
| NumWebPurchases | number of purchases made through company’s web site |
| NumWebVisitsMonth | number of visits to company’s web site in the last month |
| Recency | number of days since the last purchase |

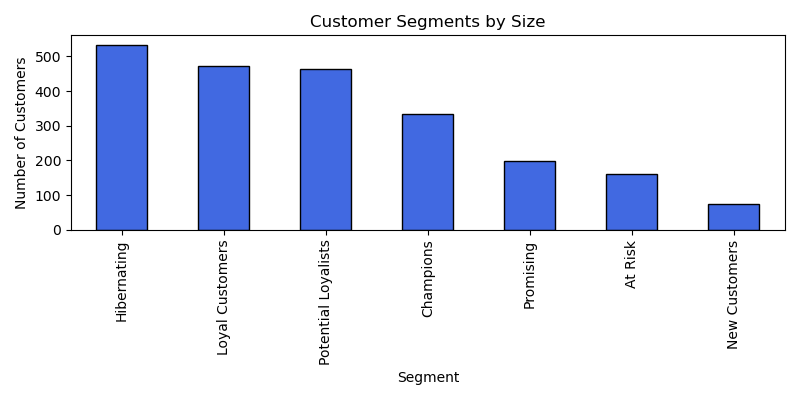
**3. Data Cleaning & Feature Engineering**

* Imputed missing income with median values
* Removed outliers and there were no duplicates
* Parsed and standardised dates
* Engineered features:
  + Age (2010 - Year\_Birth (>1935))
  + Tenure (days since Dt\_Customer)
  + Total spend (sum of product categories)

**4. RFM Segmentation**

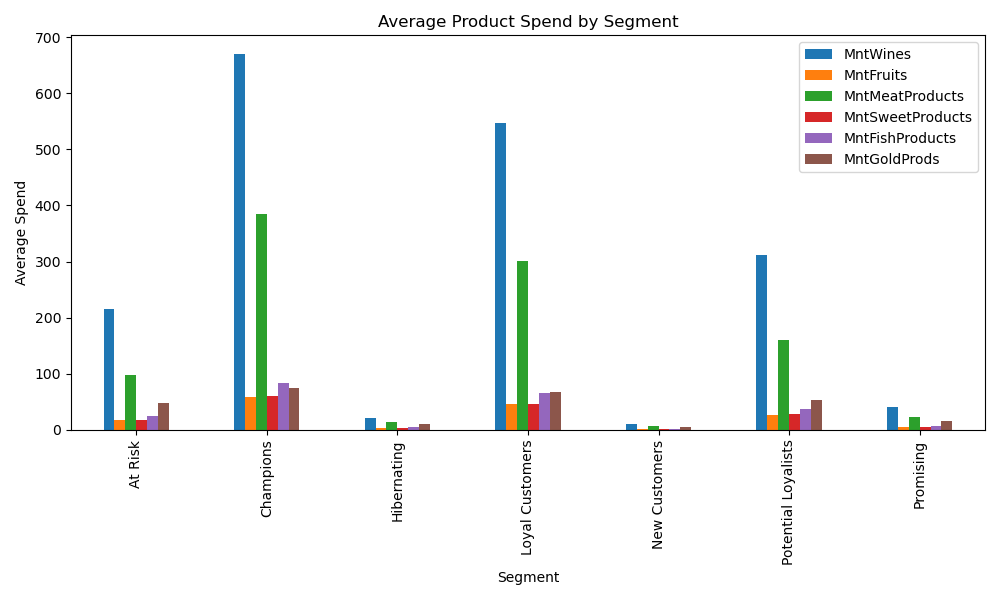
* Calculated Recency, Frequency, and Monetary (sum of purchases/channels, total spend)
* Scored each metric (quintiles, 1=lowest, 5=highest)
* Combined for RFM segment code (e.g., 555 = Champion)
* Assigned business-meaningful segment names

**Segment Distribution:**

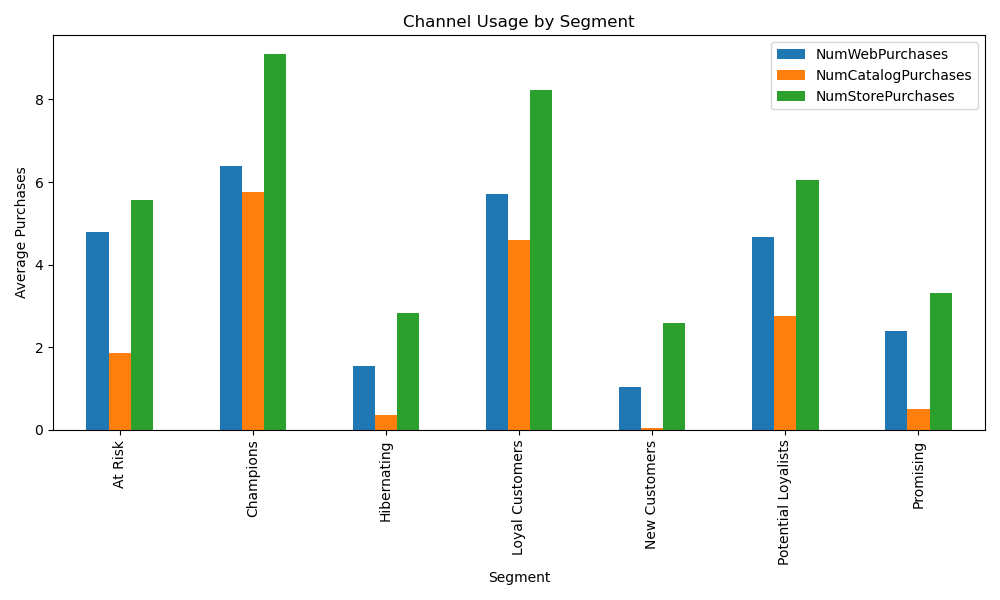


**5. Deep Profiling of Segments**

* **Demographics by segment:** Age, income, household size, education
* **Product preferences:** Spend by category per segment
* **Channel usage:** Web, catalogue, store purchase averages by segment
* **Campaign response:** Acceptance rates per segment

**Product spend by segment**

**Channel use by segment**

****

**Income by segment**

**A graph of blue bars with white text

AI-generated content may be incorrect.**

**6. Campaign Response & ROI Analysis**

* Calculated campaign acceptance rates by segment and campaign
* Estimated expected revenue by segment/campaign
* Found highest ROI for Champions and Loyal Customers (esp. Campaigns 4 & 5)
* Lower ROI for At Risk, Hibernating, and New Customers

**Campaign response by segment**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Segment | Accepted Cmp1 | Accepted Cmp2 | Accepted Cmp3 | Accepted Cmp4 | Accepted Cmp5 | Response |
| At Risk | 0.031056 | 0.006211 | 0.068323 | 0.080745 | 0.006211 | 0.055901 |
| Champions | 0.188623 | 0.029940 | 0.092814 | 0.134731 | 0.215569 | 0.332335 |
| Hibernating | 0.000000 | 0.003745 | 0.065543 | 0.007491 | 0.000000 | 0.039326 |
| Loyal Customers | 0.116773 | 0.027601 | 0.074310 | 0.140127 | 0.135881 | 0.203822 |
| New Customers | 0.000000 | 0.000000 | 0.066667 | 0.000000 | 0.000000 | 0.093333 |
| Potential Loyalists | 0.045356 | 0.008639 | 0.066955 | 0.077754 | 0.053996 | 0.140389 |
| Promising | 0.000000 | 0.000000 | 0.075377 | 0.015075 | 0.000000 | 0.125628 |

**Campaign-specific expected revenue**

|  |  |
| --- | --- |
|  | Expected Revenue (AUD) |
| AcceptedCmp1 | 87227.05 |
| AcceptedCmp2 | 18172.30 |
| AcceptedCmp3 | 98736.18 |
| AcceptedCmp4 | 101159.15 |
| AcceptedCmp5 | 98130.43 |

**Bar chart: Expected revenue by segment**

A graph with text on it

AI-generated content may be incorrect.

**7. Predictive Modelling**

* Built logistic regression model to predict campaign response (target: Response=1)
* Features: RFM scores, demographics, spend, channel usage
* Standardised all features before fitting model
* **Performance:**
  + ROC-AUC = 0.80
  + Precision (responders): 0.62
  + Recall (responders): 0.20
  + F1 (responders): 0.30
* **Top predictors:** Recency, wine spend, monetary value, web/catalogue engagement

**Classification Report Table:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | f1-score | support |
| 0 | 0.87 | 0.98 | 0.93 | 572 |
| 1 | 0.66 | 0.19 | 0.29 | 100 |
| Accuracy |  |  | 0.86 | 672 |
| Macro avg | 0.76 | 0.59 | 0.61 | 672 |
| Weighted avg | 0.84 | 0.86 | 0.83 | 672 |

The model has strong overall accuracy, but recall for responders is lower (20%), which is typical for imbalanced marketing datasets. The model is conservative in predicting responses, achieving a precision of 62% for the positive class.

**Feature Importance:**

|  |  |
| --- | --- |
| Feature | Coefficient |
| R\_Score | 0.691 |
| MntWines | 0.569 |
| M\_Score | 0.552 |
| NumWebPurchases | 0.244 |
| Kidhome | 0.237 |
| MntMeatProducts | 0.224 |
| NumCatalogPurchases | 0.141 |
| MntFruits | 0.095 |
| MntGoldProds | 0.064 |
| F\_Score | 0.025 |
| Age | 0.003 |
| MntSweetProducts | -0.044 |
| MntFishProducts | -0.076 |
| Income | -0.296 |
| Teenhome | -0.375 |
| NumStorePurchases | -0.807 |

**8. Key Insights & Recommendations**

* **Champions and Loyal Customers:** Target with premium offers and loyalty rewards. Highest ROI for Campaigns 4 & 5.
* **Potential Loyalists:** Focused nurturing (Campaigns 3 & 4)
* **At Risk, Hibernating, New:** Cost-effective, automated re-engagement. Low ROI from mass campaigns.
* **Predictive model** should be used to select best targets for new campaigns.

**9. Limitations & Further Work**

* No transaction-level timestamps, limits cohort/churn modelling
* Predictive model could be enhanced with more features or advanced algorithms (e.g., tree-based models, resampling)
* Business context assumptions: campaign costs, channel effectiveness, etc.