Contents

ACME Inc Marketing Strategy

CRISP DM model

1. Business Understanding

Build a profile of the customers who are most likely to respond to a new marketing campaign based on the prior promotions of the company. Identify a modeling node and build a stream to create a mailing list of customers to be targeted.

The evolving world of financial offerings requires modern marketing strategies that provide accurate consumer targeting. To achieve that goal ACME Inc is creating a new marketing strategy aimed at increasing the number of products each of our customers has and understand what drives a positive response from them in terms of acquiring new products and open more accounts.

Our data lake provides daily data about the transactions done by our clients, and this will be the main source of information to feed our data-driven decisions along with a capable analytics team and the use of machines learning and statistical methods.

The expected result is a list of the customers with a higher probability of answering the calls and making a purchase.

2. Data Understanding

The dataset used for creating this new marketing strategy contains 31 variables ranging from information on past customer responses to marketing campaigns, along with demographic and financial details. 18 Key fields were selected:

- **Age:** Customers age (Integer)
- **Campaign:** Identifier of campaign (1,2,3,4) (Integer)
- **Response:** 1 positive response, 0 not response (Boolean)
- **Purchase:** 1 customer purchased, 0 not purchase (Boolean)
- **Debt_equity:** ratio between the value of the total debt and the capital of a customer. (Double)
- **Gender:** Male, female (Boolean)
- **Income:** Customer income (Integer)
- **Months_current_account:** Months with the current account active (Integer)
- **Months_customer:** Number of months a being an ACME customer. (integer)
- **Call_center_contacts:** How many call attempts an agent did to each customer. Integer)
- **Loan_accounts:** How many loan accounts has each customer. (Real) has negative values.
- **Number_products:** How many products a customer has. (Integer)

- **Number_transactions:** Transactions per customer. (integer)
- **Non_worker_percentage:** Percentage of non-working customers. (Double)
- **White_collar:** Percentage of white-collar workers among customers.(Double)
- **Rfm_score:** Customer value based on Recency, Frequency, and Monetary value of purchases. (Double)

3. Data Preparation

Data preparation involved addressing missing values and inconsistencies. Specifically:

- **Missing Values:** Missing values were not identified in any of the variables selected, the procedure used for this involved counting the records of each row and after that we found that there were blank values but not null or missing values.
- **Outliers:** Outliers were examined using visualization techniques, in this the case by using the histogram we could identify that for most of the nominal variables the data distributions follow an almost normal distribution and potential outliers were investigated further for those variables that have some.
- No fields were derived from the dataset since the calculations of ratios were done previously by the data engineering team.
- In order to analyze the impact of each campaign we split the dataset in four subsets to perform a descriptive analysis and find which campaign has been more effective in terms of answered calls and product purchases.
- The same procedure was done for the variable **response** since we wanted to find why most of the customers were not responsive to previous marketing campaigns. This procedure brough us to create two subsets.
- Let´s clarify that the splitting in subsets for specific variables is just a part of the analysis and is not part of the modelling approach used to create the final output but is essential to understand what the main drivers of our customers are answering calls and purchasing new products.
- Additionally, we noticed that the variables **Debt_equity** and **Age** have the same exact values for each row which seems like an error in the data and thus we excluded this column from the analysis.

4. Modeling.

Modeling Approach:

The modeling technique used for this is a statistical method called **CHAID** which stands for Chi-squared Automatic Interaction Detection, is a statistical technique used for building decision trees in data analysis. It was chosen because its methodology allows to perform customer segmentation with ease and in this case will help us determine the main drivers or predictors that lead to a customer purchase.

Model Building:

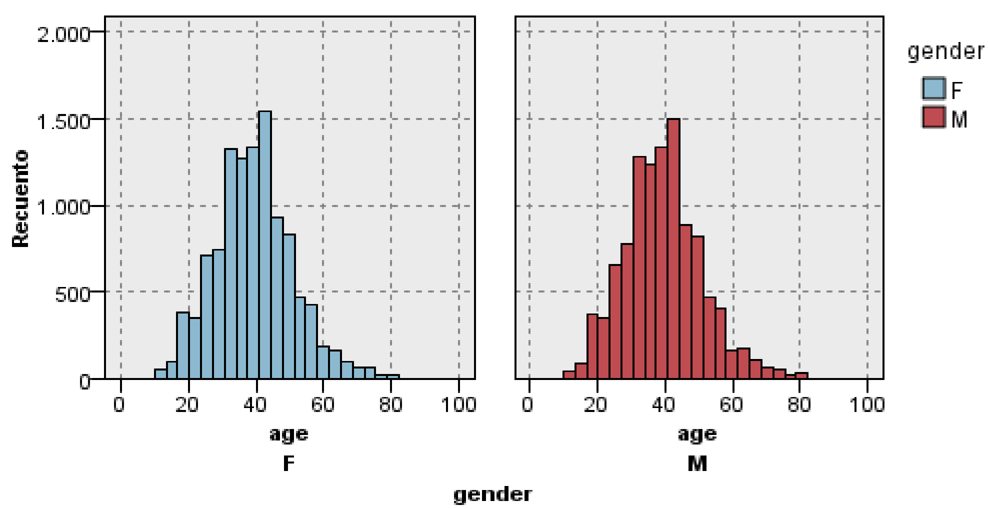
1. **Data Partitioning:** The data was split into training and testing sets using the This allows evaluating the model's performance on unseen data.

Results:
The original dataset contains 21.927 records where we identified that the customers that responded positively to one or more campaigns corresponds to 10.61% of the sample which accounts for 2.328 records and those who did not responded to any of the marketing campaigns represent 89,38% of the sample totalling 19.599 customers. From those customers who responded to any of the marketing campaigns (2.328) 69% of them (1602) bought a product and 31% (726) did not.

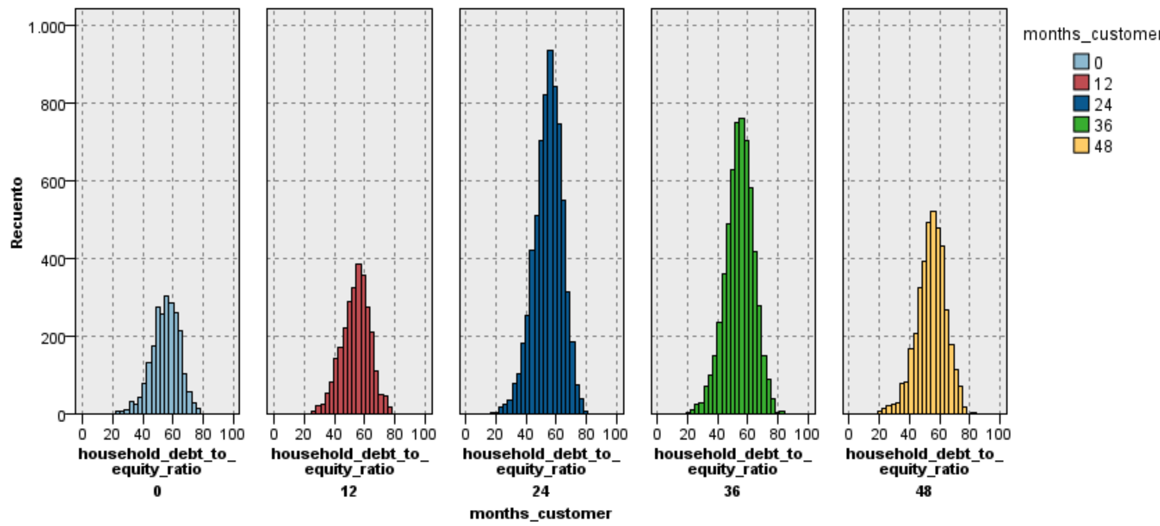
When it comes to gender, the customer base is balanced between both genders where the males represent 49.6% (10.843) of the sample and females 50.4% (11.084).

Let´s take a look at the distributions of the principal nominal variables:

Age: the average age of the customers is 39 years and when we look at the histogram, we can see that it follows a normal distribution

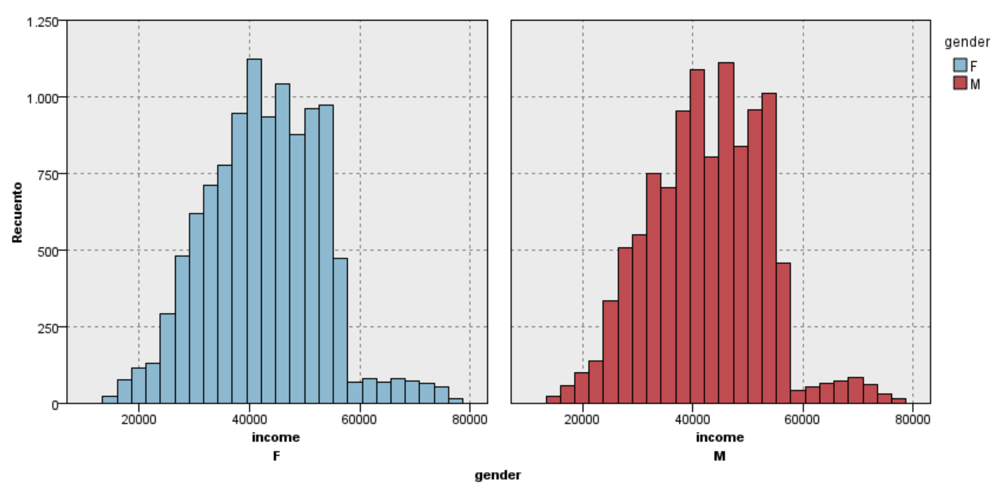


Household debt to equity ratio: The average debt to equity for the sample is 54.53% which indicates how much a customer has in debt in proportion to the total value of its capital. It follows an almost normal distribution too.



This chart shows the household debt-to-equity ratio for customers categorized by their enrollment duration at ACME Inc. Customers enrolled for 24 to 36 months have the highest representation within the total customer base.

Income: the average income of our customers base is $42.680 and when we take a look at the graphic, we can observe that those earning more than $60.000 and more represent a small part of the sample regardless of the gender. Additionally, it follows an almost normal distribution but in this case the bell shape is not as clear as in the other variables studied previously.

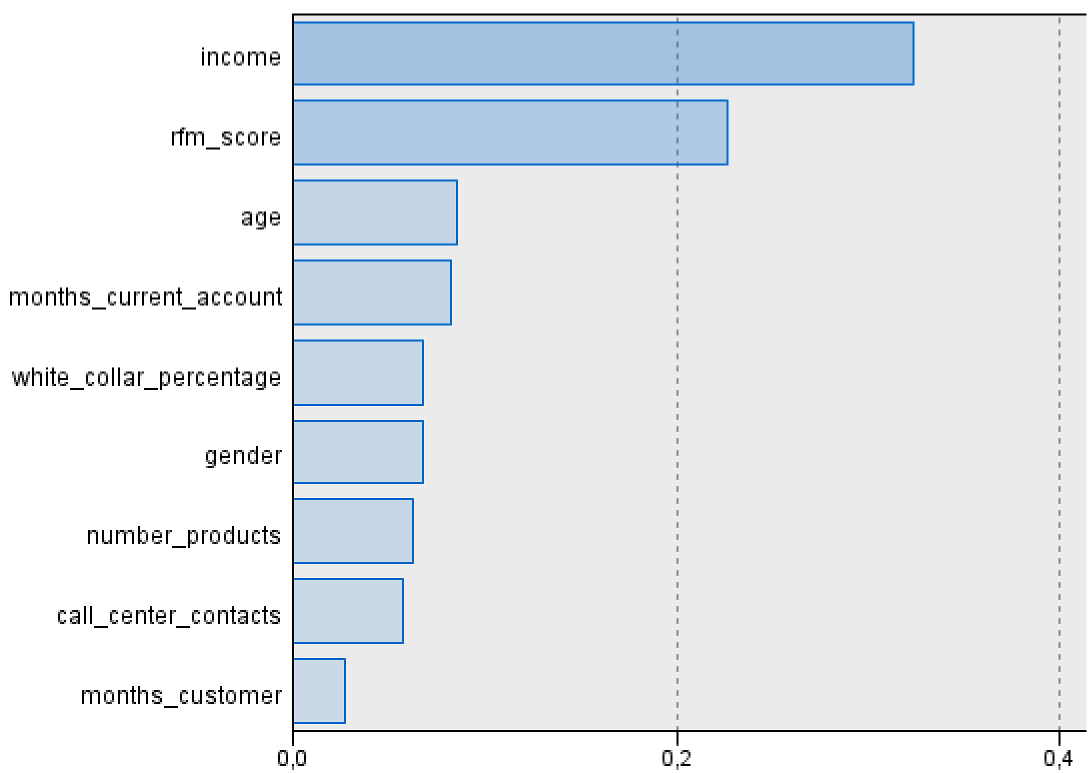


Finally for the variables of **non worker percentage (NWP)** and **white-collar percentage (WCP)** we observe that in average the non worker customers represent 10,9% of the sample and the white-collar workers represent in average 33% of the sample which is a positive indicator since is preferable to have customers with stable jobs and higher earnings than customers with a riskier financial situation.

After examining some of the most relevant nominal variables along with their distribution, we proceed to execute the **CHAID** model mentioned before in order to identify profiles of customers most likely to respond to future marketing campaigns.

The results show that the main drivers for a customer to make a purchase are the following variables in order:

Variable	**Importance in percentage**
1) Income	32%
2) Rfm_score	23%
3) Age	9%
4) Months_current_account	8%
5) White_collar_percentage	7%
6) Gender	7%
7) Number_products	6%
8) Call_center_contacts	6%
9) Months_customer	3%



5. Evaluation

- Income emerged as the most significant factor influencing purchase decisions, highlighting the importance of targeting campaigns based on income levels.
- RFM Score played a substantial role, demonstrating its effectiveness in identifying valuable customer segments.
- While Age showed a moderate impact, it can still be used for further segmentation within income groups.
- Months with Current Account suggests loyalty might increase with account tenure.
- White-Collar Percentage potentially indicates spending capacity or professional needs influencing purchase decisions.

6. Recommendations:

Based on these findings, the following recommendations are suggested:

1. Develop Targeted Campaigns: Design marketing campaigns tailored to specific income brackets, incorporating RFM scores for further refinement.
2. Refine Segmentation: Explore segmenting customers based on combinations of these key variables (e.g., high-income, young professionals) for more precise targeting.
3. Leverage RFM Score: Utilize RFM scores to prioritize high-value customers for retention efforts and identify potential churn risks.
4. Monitor and Adapt: Continuously monitor campaign performance and adjust strategies based on results.

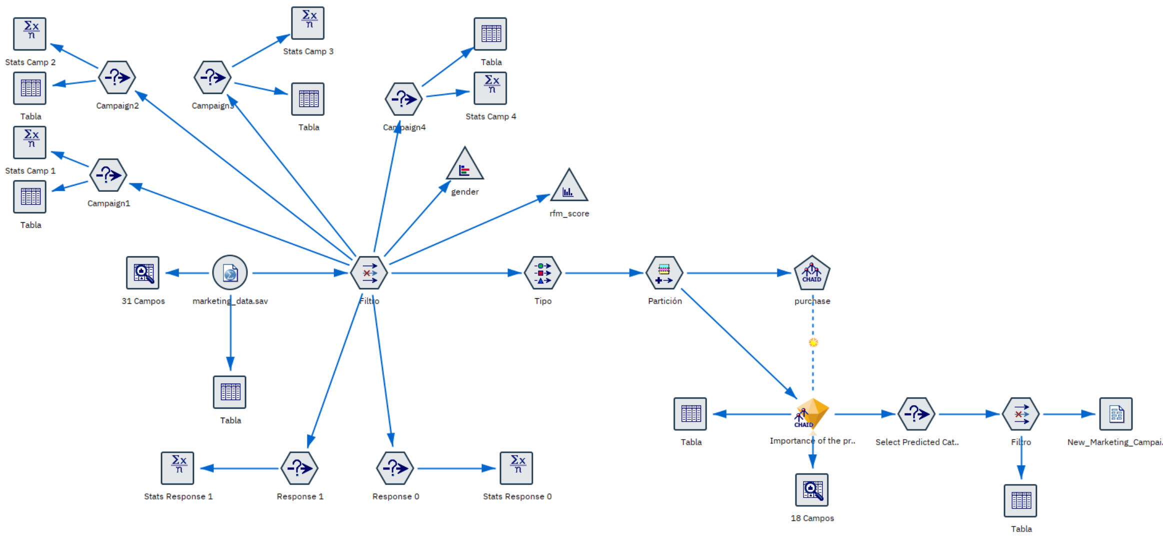
Next Steps:

- Develop detailed customer profiles based on the identified segments.
- Create targeted marketing campaigns based on the identified drivers and customer segments.
- Implement A/B testing to optimize campaign messaging and creative for different segments.
- Track campaign performance and adapt strategies based on results.

Conclusion:

The CHAID model successfully identified key factors influencing customer purchase behavior. These insights can be used to develop targeted marketing campaigns with a higher likelihood of customer response and improved ROI. By continually monitoring results and adapting strategies, the effectiveness of marketing efforts can be further enhanced. The final output provided a list with 980 records with levels of confidence 58% for 445 customers and 72% for 535 customers which doesn´t seem low considering the uncertainty behind factors such as the probability of a customer answering a call and other aspects that not necessarily depend on the agents actions.

7. Appendix A.


Stats Camp 2
Tabla
Campaign2
Stats Camp 1
Tabla
Campaign1
Stats Camp 3
Tabla
Tabla
Stats Camp 4
Campaign4
gender
rfm_score
31 Campos
marketing_data.sav
Tabla
Filtro
Tipo
Partición
purchase
Tabla
Importance of the pr..
Select Predicted Cat..
Filtro
New_Marketing_Campai..
18 Campos
Tabla
Stats Response 1
Response 1
Response 0
Stats Response 0