**Introduction: Customer Insights to Drive Future Business for Financial Organizations**

In today's competitive and rapidly evolving landscape of financial services, data has emerged as the lifeblood of successful operations. Financial firms, ranging from banking institutions to investment companies, have recognized the immense value of tracking certain customer information for future business opportunities. This data-driven approach is not merely an exercise in data collection but a strategic imperative that empowers firms to understand their customers better, tailor personalized solutions, mitigate risks, and ultimately drive growth in a customer-centric marketplace.

First and foremost, tracking customer information grants financial firms an unparalleled insight into the needs, preferences, and behaviors of their clients. By gathering and analyzing data on spending patterns, investment choices, and financial goals, firms can construct comprehensive customer profiles. Armed with this knowledge, they can craft highly targeted and personalized services, products, and recommendations. Whether it's offering tailored investment portfolios or designing bespoke banking packages, these tailored offerings enhance customer satisfaction, foster long-term loyalty, and attract new business opportunities.

Secondly, tracking customer information is a vital component in managing and minimizing risks within the financial industry. Analyzing historical data and financial behavior allows firms to identify potential red flags and detect early signs of fraudulent activities. By staying vigilant in monitoring customer accounts and transaction patterns, financial institutions can prevent losses due to fraudulent schemes and protect both their customers and their own reputations.

Thirdly, the practice of tracking customer information enables financial firms to embrace a proactive approach to anticipate and respond to market trends. Armed with comprehensive data analytics, these firms can identify emerging opportunities and potential threats well in advance. This invaluable foresight empowers them to adapt their offerings, fine-tune their strategies, and stay ahead of competitors. As customer needs evolve and new market niches emerge, these agile firms can quickly seize opportunities to expand their portfolio and strengthen their market position.

Lastly, tracking customer information fosters a culture of continuous improvement within financial firms. Utilizing data-driven insights, firms can evaluate the success of their strategies and adapt accordingly. Identifying areas for improvement and optimizing internal processes can lead to greater efficiency, cost reduction, and enhanced customer service. By harnessing the power of customer data, financial firms can foster a learning organization that thrives on innovation and delivers exceptional value to its clients.

**Analysis and Models**

This section will provide additional information in the data that was used to complete this study, the data preparation & cleaning process, and an overview of the model used to conduct additional analysis.

**About the Data & Business Scenario:**

For this study, the team was hired by a marketing department of a financial firm that keeps records on customers, including demographic information and, number of type of accounts. When launching a new product, such as a "Personal Equity Plan" (PEP), a direct mail piece, advertising the product, is sent to existing customers, and a record kept as to whether that customer responded and bought the product. Based on this store of prior experience, the managers decide to use data mining techniques to build customer profile models*.* The overall purpose of this study is to determine the top Association Rules to help with Pep Marketing Campaign. The data is organized as follows:

*Figure 1 – Table explaining raw dataset attribute names and datatypes.*

A close up of a list

Description automatically generated

**Data Preparation & Cleaning:**

Before conducting any form of analysis, the data must be formatted in a way that allows for calculations and analysis to occur. The team imported the csv file into a data analysis platform and inspected how many observations and attributes are in the dataset. Looking at the raw data below, we can see that there are 600 rows and 12 columns:

*Figure 2 – “Bank Data” dataset structure*

A screenshot of a computer code

Description automatically generated

From the picture above, the team assessed that datatype conversion is necessary prior to conducting any form of analysis. The team converted all the datatypes to either a factor datatype to help with the categorical values. The overall purpose of datatype reassignment is to prepare the data for aggregation or calculations. Ensuring that the attributes are in a suitable format will can lead to more accurate results.

*Figure 3 – Table highlighting converted datatypes for each attribute.*

A screenshot of a computer program

Description automatically generated

After reassigning the datatypes for each of the 12 attributes, the team viewed the newly formatted data to get preliminary insight on the dataset.

*Figure 4 – View of formatted data*

A screenshot of a computer

Description automatically generated

**Null or Missing Values:**

There are no missing values in this dataset as shown in the figure below:

*Figure 5 – Confirmation of no NULL values in each of the columns*



**Baseline Information:**

Before conducting any form of analysis, the team wanted to see how data is dispersed across the attributes that have numeric datatypes: Age, Income and Children. Based on the figure below, we can see the summary statistics as well as the visualization that confirms the summary statistic output.

*Figure 6 – Summary Statistics for Numerical Attributes and their Visualizations*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Min | 1st Quartile | Median | Mean | 3rd Quartile | Max |
| Age | 18.00 | 30.00 | 42.00 | 42.40 | 55.25 | 67.00 |
| Income | 5014 | 17264 | 24925 | 27524 | 36173 | 63130 |
| Children | 0 | 0 | 1 | 1.012 | 2 | 3 |

A graph of age distribution

Description automatically generated

A graph of income distribution

Description automatically generated

A graph of a number of children

Description automatically generated

**Results**

First, the team decided it was best to analyze each of the attributes to gain foundational understanding of each of the attributes. The team created bar plots to help visualize the behavior of 11 attributes broken down by their categorical values.

*Figure 7 – Bar plots for the 12 attributes*

A graph of a number of people

Description automatically generatedA graph with a couple of squares

Description automatically generated

A graph of a number of bars

Description automatically generatedA graph of a number of bars

Description automatically generated

A graph with a number of squares

Description automatically generatedA graph of children distribution

Description automatically generated

A graph showing a number of car ownership

Description automatically generatedA graph showing a number of savings account

Description automatically generated

A graph showing a number of ownership

Description automatically generatedA graph showing a number of ownership distribution

Description automatically generated

A graph showing a number of gray squares

Description automatically generated

**Attribute Breakdown Results**

Based on the output above, the team derived the following observations:

1. Majority of the bank members reside within an urban or inner-city environment.
2. Most of the bank members fall within a low-income bracket.
3. Most of the bank members have families with no children.
4. Since most of the bank members are in an urban environment, approximately half of the members do not own cars and most likely rely on public transportation.
5. Most of the bank members of this financial institution do not own homes or have a mortgage loan with the bank.
6. Most of the bank members do not own a “PEP” account (newest bank product/service).

**Association Rule Mining**

The team conducted three trials of running the apriori algorithm on the bank data to determine the top rules that can predict a “pep” value of Yes. Here are the parameter values for each of the trials:

*Figure 8 – Apriori Algorithm Support and Confidence Values per trial*

|  |  |  |
| --- | --- | --- |
|  | *Support Value* | *Confidence Value* |
| *Trial 1* | *0.002* | *0.5* |
| *Trial 2* | *0.015* | *0.8* |
| *Trial 3* | *0.021* | *0.91* |

Each of the trials produced different rules with different item sets. It is also worth noting that Trial 1 did not produce any rules that had a ‘pep’ value of yes on the right-hand side.

**Trail 1 Results**

Trial 1 created over 74,000 rules in about .23 seconds. What was interesting with Trial 1’s results was that majority of the item set values had a right hand side value of either “Age = TEENAGER”, “Region = “Suburban” or “Children = 3”. The team assessed that based on the data provided, the apriori algorithm using specific parameter values influenced the output.

*Figure 9 – Trial 1 Results (rulesPep1)*

*A white background with black lines

Description automatically generated*

**Trial 2 Results**

For this trial, the team updated the support and confidence values and came up 13,087 rules. This is a significant decrease from the rules result from trial 1. This confirms that finding the correct support and confidence values is critical in trying to obtain accurate item set rules. This trail produced a wider spread of right-hand side options. Of the 100 pages of itemset rules, the team had to filter to page 50 of the results to get what we were looking for.

For the rules that had a “Pep” value of “Yes” here are the pertinent item sets:

*Figure 10 – Trial 2 rules output*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| LHS Item Sets | Support Value | Confidence | Lift | Count |
| Income=45,000+, children= 1 | 0.0267 | 1 | 2.189781 | 16 |
| Age=Forties, children=1 | 0.060 | 1 | 2.189781 | 36 |
| Age = Sixties, Income= 4500+, children = 1 | 0.020 | 1 | 2.189781 | 12 |
| Sex = Male, Income= 45000+, children = 2 | 0.020 | 1 | 2.189781 | 12 |
| Income = 45000+, children = 2, car = No | 0.015 | 1 | 2.189781 | 9 |



The table showcase strong relationships between different items in the dataset. The high confidence and lift values indicate that the presence of items on the left-hand side of the rules significantly influences the occurrence of items on the right-hand side, and for this case, “Pep”. These rules highlight that there is specific income bracket, children amount, age and car ownership aspect to ensuring a “Pep” product is bought.

**Trial 3 Results**

For trial 3, the team updated the support and confidence values again and was only given 1335 rules. This behavior highlights that the setting higher support and confidence thresholds in Association Rule Mining filters out less frequent and weaker rules. This process ensures that the output contains only the most relevant and significant rules, making it easier for analysts and stakeholders to interpret and act upon the discovered associations. It also helps avoid overwhelming results with numerous less meaningful rules that might not be practical or valuable for decision-making in real-world applications.

Similar to trial 2, the trial returned similar item set values to accurately predict a “Pep” value of Yes.

*Figure 11 – Trial 3 rules output*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| LHS Item Sets | Support Value | Confidence | Lift | Count |
| Income=45,000+, children= 1 | 0.0267 | 1 | 2.189781 | 16 |
| Age=Forties, children=1 | 0.060 | 1 | 2.189781 | 36 |
| Income= 4500+, children = 2, mortgage = No | 0.021 | 1 | 2.189781 | 13 |
| Income= 45000+, children = 1, Save\_Act = Yes | 0.0267 | 1 | 2.189781 | 16 |
| Income = 45000+, children = No, Married = No | 0.031 | 1 | 2.189781 | 19 |

A screenshot of a computer

Description automatically generated

The interesting behavior in this trial is that it produced the same Lift value and generally the same support values from trial 2. However, the item sets and frequency of the item sets differ. For this trial, the rules produced item sets that include the attributes of “Save\_Act”, and “Mortgage”.

**Conclusions**

In conclusion, the practice of tracking customer information is not merely an administrative chore for financial firms but a strategic undertaking with far-reaching implications. By harnessing data analytics and leveraging customer insights, these firms can cultivate personalized services, mitigate risks, embrace market opportunities, and foster a culture of continual improvement. In a business landscape where competition intensifies and customer expectations soar, the integration of data-driven decision-making becomes an essential ingredient for sustainable success and future growth in the financial services industry.

The observations made during the analysis offer important implications for the bank’s marketing team to determine who to create “Pep” promotional emails for. Based off the output, marketing and business development teams should target bank members that have an income over 45,000, a low number of children, non-homeowners, and members that are within their 40’s. Based off the association rules, these are the attributes that is likely to result in the purchase of a PEP product.

By tracking customer information, financial institutions fosters a culture of continuous improvement with its customers as well as provides the opportunity to identify key populations within their pool of members. Utilizing data-driven insights, firms can evaluate the success of their strategies and adapt accordingly. Identifying areas for improvement and optimizing internal processes can lead to greater efficiency, cost reduction, and enhanced customer service. By harnessing the power of customer data, financial firms can foster a learning organization that thrives on innovation and delivers exceptional value to its clients.