Executive Summary:

This report analyses a random transactional dataset of 54,785 records. The analysis identified six unique customer segments to create tailored marketing strategies to cross-sell and deep-sell financial products.

The data showed that customers are across different age groups: below 18 (6.42%) and above 80 (5.57%). Customers aged 80 or higher were excluded from the dataset due to their perceived minimal movement in product ownership. 42.25% had zero income with the bank, highlighting the need for strategic efforts to incentivise these customers to transfer their income to the bank. Customers with monthly income of $10,000 or more were considered outliers. Offering them exclusive banking services, such as private banking, will be more beneficial in promoting high-value products like mortgages or investment trusts.

The largest segment, Low Income – No Product Owners, comprises 59.0% of the customer base. The smallest segment, High-Income Homeowners, includes 2.5% of the customer base and comprises mortgage holders with limited other product ownership.

Strategic recommendations for each segment were developed based on their unique profiles. For the **Low Income – No Product Owners** segment, introducing credit facilities and incentives like first-time product engagement through promotions and robust onboarding processes is recommended. For **Moderate Income – Savers**, promoting low-risk investment funds, personal loans, auto loans, and mid-tier credit cards with favourable terms would be beneficial. **High Income – Credit Heavy Savers** should be targeted with exclusive credit card upgrades, mortgages, and investment trusts tailored to their financial goals. A rewards program to encourage additional product ownership is suggested for **High Income – Investors**. Finally, for the **High Income – Debtors and Home Owners segments**, offering dual product ownership programs, optimising costs with special interest rates, and promoting high-tier credit cards can enhance product engagement and customer satisfaction.

This comprehensive segmentation and strategic analysis provide a foundation for targeted marketing efforts to enhance product ownership, customer satisfaction, and overall financial inclusion.

Table of Contents

Statistical Techniques and Variables:

Data Description:

A random transactional dataset of 54,785 records with seven binomial and three metric variables. The data consists of 2 metric **demographic variables**, Age in years, Income in Dollars, and eight **domain-specific variables**. Within the domain-specific variables are seven binomial variables: Churn, Loan, Credit Card, Checking Account, Mortgage, Savings Account, Investment Trust, and one metric variable: Number of Transactions.

Data Cleaning:

Descriptive statistics – frequencies were used to check for missing values for each variable and information on central tendency, range, minimum, and minimum. **Figure 1.0** in the appendix summarises this information.

While analysing age, the customer demographic is divided into three broad categories. 3,516 - 6.42% of the customers are below the age of 18, customers between the age of 18 to 80 are 48,217 – 88.01%, whereas 3,052 – 5.57% of customers with the age of 80 or higher. According to Stats NZ (2023) life expectancy review, the average life span of males and females is 80 and 83, respectively. With gender data not available, customers aged 80 or higher were removed from the data as they are considered fully matured clients with little to no movement in future product ownership. Customers below 18 are considered future clients, with moderate variation in product ownership; they play an integral part in strategy formulation. **Figure 1.1** describes these demographic categories across product ownership.

Frequency tables with cumulative frequency were created for all metric variables to summarise pool size, The table for Income indicated 42.25% - 23,145 customers with zero Income with the bank. Further analysis showed that these customers were spread across multiple product ownership; thus, it was essential to keep them in the dataset to develop incentivising strategies to encourage customers to transfer income into the bank. Customers with a monthly income of $10,000 or higher were considered as “high earning individuals”; this constitutes about 0.16% - 86 individuals; they were removed from the data set because these customers should be segmented based on income alone and be targeted separately in order cross-sell and deep sell high-value financial products. Analysis of these individuals showed that these 86 customers were spread across multiple product ownerships. **Figure 1.2** summarises income distribution and product ownership.

The frequency table for churn, **Figure 1.3,** shows that 1,371 – 2.5% of the customers had left the business. According to Verhoef et al. (2003), it is vital to exclude churned customers so that marketing efforts are targeted at those most likely to respond to cross-sell and deep-sell initiatives. This targeted approach is more efficient and can lead to better investment returns.

Statistical Techniques and Variable Selection:

5 binomial domain-specific variables, Loan, Credit Card, Mortgage, Savings Account, Investment Trust, and one metric demographic variable, Income, were taken for the cluster analysis.

Checking Account was not chosen for this analysis because the frequency table showed 100% of the sample size to own it, with only 24 customers without it. **Figure 1.4** shows each product and its percentage ownership across customers. Clustering is based on variability, supported by Hair et al. (2019), which states that a variable that is the same across all customers does not help distinguish between groups. Thus, this variable was dropped.

A two-tailed Pearson correlation test was conducted between all metric variables. The test showed **Income and Number of Transactions** to be moderately correlated with a positive Pearson correlation score of 0.744 and a significance of less than 5%. **Age and Income** were also significant ($p<0.05$); the Pearson correlation score of 0.205 suggested a weak correlation between the two variables. **Number of Transactions and Age** had a Pearson correlation score of 0.146 and a significance of less than 5%. This showed that there was a very low correlation between the two variables. Refer to **Figure 1.5** for the correlation test. Income and number of transactions were initially selected for clustering. The silhouette score of the cluster was 0.6, which was lower than when compared to clustering without Number of Transactions. The addition of Number of Transactions diluted the clusters, and hence, it was dropped. **Figure 1.7** shows the cluster results with and without the variable.

Variables were also chosen based on a trial-and-error approach. Hair et al. (2019) support this approach, stating that refining the variable set based on preliminary results improves cluster quality. The final variables were chosen as they provided cluster variability, a silhouette score of more than 0.5, and catered to the strategic goal of targeting customers based on product ownership. **Figure 2.1** shows multiple cluster analyses performed.

According to Shih et al. (2010), a two-step cluster analysis is better suited for cluster segmentations for larger datasets containing continuous and categorical variables. Since the data used had metric and non-metric variables, the log-likelihood distance was used. One-way ANOVA was performed to check the significance ($P<0.05$) of cluster membership with metric variables, and the Chi-Square test was conducted to evaluate the significance ($P<0.05$) with the non-metric variables. Milligan and Cooper (1987) describe these tests as validating the comprehensiveness of the clustering performed.

Evaluation, Segment Profile and Labels:

Evaluation:

The two-step cluster analysis was forced to create six unique segments. Clusters had a silhouette score of 0.8. Refer to **Figure 1.9** to see cluster results. In the paper, Arbelaitz et al. (2013) suggest that a score higher than 0.5 indicates that each cluster is separated and cohesive. While analysing, each cluster showed at least one complete product ownership. The high variability of individual clusters supported the strategic goal of segmenting customers into six unique clusters. In the paper, Dolnicar & Grün (2008) suggest that when automatic clustering generates inferior solutions, setting a predefined number of clusters can create more structure in the data.

The clusters are ordered based on size. The largest cluster is the Low-Income No Product Owners, with 32,263 (59.0%) customers, and the smallest cluster is High-Income Home-Owners, with 1,376 (2.5%) customers. See **Figure 1.6** for a tabular representation of the 6 clusters.

One-way ANOVA's test of homogeneity of variances with income resulted in the Levene statistic being significant ($P<0.05$). This meant that equal variances were not assumed. The Welch test showed that the generated cluster membership is significant ($P<0.05$) compared to income. Tamhane's T2 test resulted in significantly different mean values. **Figure 1.8** shows ANOVA results.

Pearson Chi-square test showed significance ($P<0.05$) across all non-metric variables selected. This meant that each cluster was unique with regard to product ownership. **Figure 2.0** shows the results for each non-metric variable.

The analysis was run multiple times to optimise the number of clusters. Silhouette score, variability, and customer segmentation based on the strategic goals were analysed. **Figure 2.2** shows two out of multiple cluster analyses performed.

Segment Profiles and Labels:

Below are each cluster's characteristics.

Cluster 1: Low Income – No Product Owners: Comprises 32,263 individuals, accounting for 59.0% of the total sample. This cluster exhibits a profile characterised by a complete absence of ownership in all financial instruments. With an average monthly income of $357.46, this cluster reflects a population with relatively low earning capacity or irregular income streams.

Cluster 5: Moderate Income – Savers: This cluster represents 21.0% of the sample size with 11,514 individuals. Every member in this cluster does not own a credit card, investment trust, loan, or mortgage. Despite this, they all own savings accounts, exhibiting a low-risk approach towards saving money. With an average income of $937.85, this cluster suggests a moderate income level. This profile implies a conservative approach towards savings with no interest in credit or investment opportunities.

Cluster 2: High Income – Credit Heavy Savers: This cluster, comprising 4,687 individuals, accounts for 8.6% of the total sample. The vast majority, 95.1%, own credit cards, indicating a reliance on credit for transactions or potentially building credit history. With an average income of $1,673.67, members of this cluster seem to have high earning capacity. Despite their high income, none own investment trusts, loans, or mortgages, suggesting a cautious approach towards borrowing and investment activities, or these customers own products with a different bank. Only 49.7% of individuals in this cluster own savings accounts. This profile suggests a group that relies on credit for transactions but may benefit from enhancing their savings habits to complement their income and credit management.

Cluster 3: High Income – Investors: Representing 4.7% - 2,583 individuals. 70.70% do not own credit cards despite their high income. 100% of the cluster members own investment trusts, indicating a preference for investment vehicles beyond traditional banking products. With an average income of $1,524.51, individuals in this cluster exhibit high earning capacity. Moreover, the vast majority, 95.8%, do not have loans, and 93.0% do not have mortgages, highlighting a conservative approach to debt management and homeownership. However, 67.5% own savings accounts, indicating a recognition of the importance of savings. This cluster profile suggests a group prioritising investment and savings over debt.

Cluster 6: High Income – Debtors: Cluster size of 4.2% - 2,276 individuals. A blend of customers with credit activity and debt characterises them. A significant majority, 66.4%, of individuals do not own credit cards; members' average income is $1,469.52, indicating a high earning capacity. 100% of the individuals do not own investment trusts, indicating a lack of participation in investment opportunities beyond savings accounts; 40% own savings accounts. Interestingly, 100% of the cluster owns a loan, suggesting a reliance on borrowing for financial needs. However, none possess mortgages, indicating a preference for other forms of borrowing.

Cluster 4: High Income – Home Owners: Comprising 1,376 individuals, this cluster represents 2.5% of the sample analysed. This is the smallest cluster, with a high average income of $1457.90. The debt classification of these customers shows that 100% own a mortgage, and 86.0% have no loans, which shows that they limit their financial burden by opting for one borrowing instrument. Investing activities of these customers show that 51.2% do not own savings accounts. Even though some hold low-risk savings accounts, 96.9% do not own an investment trust.

Strategies and Recommendations:

Cluster 1: Low Income – No Product Owners: This is the largest pool of customers segmented together. With lower income and no financial product ownership, the bank's first goal should be to provide credit facility. Offering credit cards with no fee for the first year and discounts across retail outlets would encourage customers to lower their barriers towards a new financial product. Promote incentives for first-time product engagement, such as cashback schemes for using their limit on their new credit card and interest rate promotions on savings account for the first three months, according to Demirguc-Kunt et al. (2018), which states that incentives build trust and increases financial inclusion within customers. The banking industry relies on service quality; according to Keiningham et al. (2007), customer satisfaction significantly impacts customer loyalty and retention; thus, it is critical to devise a robust onboarding process while targeting customers. Surveying customers to rate their banking experience on banking products, service quality, digital experience, and customer support would identify gaps. Improving stats would help increase customer value.

Cluster 5: Moderate Income – Savers: Individuals in this cluster have a stable income with the bank, which allows them to maintain a savings account. However, the absence of investment trust suggests that they are risk-averse. Promoting low-risk investment funds with no front load, low opening balance, and no minimum balance requirements would aid in reducing the barriers towards converting these customers towards investment accounts. A customer survey to assess the risk profile and investment objective would help create a more tailored strategy. People with moderate income might be unable to afford a mortgage; thus, offering small loans such as personal loans for education or home improvement should be targeted. Auto-loans with low downpayment requirements, favourable interest rates, flexible repayment options, and lower bank charges on early full repayment would encourage customers to opt in. Advertising hedging would help the bank promote two opposite financial products simultaneously, ultimately increasing product ownership. Introduce mid-tier credit cards since they have the income to support it. Cards with lower annual percentage rates can benefit individuals in this cluster as they might have occasional large or unexpected expenses. Advocating advantages like lower accrued interest and better financial planning with the option to spend large sums.

Cluster 2: High Income – Credit Heavy Savers: Customers in this segment have the highest average income with the bank. Create targeted credit card strategies to convert the remaining 4.9% of people and offer upgraded exclusive cards with better limits, rewards and benefits. Yi and Kim (2020) suggest that personalised credit card upgrades significantly increase customer satisfaction. There is presence of ownership of savings account, 49.7% of people own it, advocating the remaining 50.3% through exclusive banking services like bonus interest rates for the first three months, and dedicated relationship manager would appeal to these individuals. Since these customers are high-earning individuals, they should also be targeted with mortgages. With 100% of the customers without a mortgage, advertising better interest rates and educating them about various types of mortgage plans specifically tailored to their needs might compel them to move their mortgage from a different bank to ours. Communicating with these customers for their financial goals through in-app surveys, notification prompts, or emails can help the bank understand what product to cross-sell. If the goal is wealth generation, the bank should promote investment trusts with high returns. With low front load and initial deposit requirements, the bank can compel them to shift their income to an investment trust. Promoting the tax benefits of having an investment trust to high-income owners can also push them to open this account. On the other hand, if the goal is to spend but the customer is low on funds, the bank can push loans as a product with competitive interest rates, less processing time and minimal documentation requirements.

Cluster 3: High Income – Investors: Customers in this segment are diversified. Ownership level is low, but it indicates a financial behaviour towards these products. A strategy to initiate a rewards program would help the bank cross-sell multiple products. Aksoy et al. (2008) found that rewards programs positively impact customer satisfaction in the banking industry. The reward program can be divided into point-based and tiered rewards. Point-based customers get points based on interaction with the owned product. Accumulated points can be redeemed for banking fees, travel vouchers, or cash back. Tiered rewards based on ownership of multiple products. Since 100% own investment trust, and 71% do not own credit cards, they can be targeted to apply for one. 7% of customers own a mortgage, and 4.2% have a loan. This suggests that customers in this segment have weak financial behaviour towards these products; thus, targeting them could increase their chances of converting into product owners. Tiered rewards can be - priority customer service, travel vouchers or discounts on banking fees. A bonus offer within the rewards program could be initiated where customers who own all products can access exclusive airport lounges worldwide or discounts on premium tickets at leading airlines.

Cluster 6: High Income – Debtors and Cluster 4: High Income – Homeowners:

Initiate a dual product ownership program where owning one product provides special offers on another. Opting for a mortgage with the bank allows customers to get an investment trust where no handling fee is charged for the first year. No minimum requirement on balance, and the investment trust can be opened with minimum documentations. Benefits such as cost optimisation can be campaigned, such as if a customer owns a mortgage and a loan, and the bank provides special lower interest rates, compelling customers to opt for this program. A top-tier credit card with exclusive discounts, low markup and priority delivery can be coupled with debt owners, promoting increased credit facility to facilitate spending. Credit card owners can be pushed to open a savings account where the bank will not charge a transaction fee if they pay bills using the account for at least six months. This strategy will compel customers to maintain balance and even incentivise the income transfer into the account. Target people with loans and offer them credit cards with low annual fees, increasing their monthly spending power.

Appendix

Figure 1: Descriptive Statistics

Statistics

		Loan	Credit card	Checkings Acc	Inv Trust	Mortgage	Savings	Income	Number transactions	Age	churn_t1
N	Valid	54785	54785	54785	54785	54785	54785	54785	54785	54785	54785
	Missing	0	0	0	0	0	0	0	0	0	0
Mean		.05	.12	1.00	.05	.03	.31	747.03	165.37	45.03	.0250
Median		.00	.00	1.00	.00	.00	.00	454.00	135.00	42.00	.0000
Mode		0	0	1	0	0	0	0	0	50	.00
Std. Deviation		.212	.320	.020	.213	.166	.464	1259.945	157.171	50.606	.15620
Variance		.045	.102	.000	.045	.028	.216	1587461.767	24702.879	2560.924	.024
Range		1	1	1	1	1	1	127399	1631	9999	1.00
Minimum		0	0	0	0	0	0	0	0	0	.00
Maximum		1	1	1	1	1	1	127399	1631	9999	1.00

Figure 1.1: Age segmentation against product ownership

Age Segmentation vs Product Ownership

Age Clusters	Count	% contribution	Loan	Credit Card	Checking Account	Investment Trust	Mortgage	Savings Account
Future Boomers	3,516	6%	-	-	3,516	21	-	184
Fully Matured	3,052	88%	2	25	3,052	75	10	1,006
Rest	48,217	6%	2,582	6,324	48,194	2,510	1,551	16,029
Total	54,785	100%	2,584	6,349	54,762	2,606	1,561	17,219

Note: *Future boomers – Below 18. Fully matured – above 80. Rest – between 18 and 80.*

Figure 1.2: Income distribution against product ownership

Income Distribution vs Product Ownership

Income Distribution	Count	Percentage Contribution	Loan	Credit Card	Checking Account	Investment Trust	Mortgage	Savings Account
No Income	23,145	42.25%	261	726	23,122	454	193	2,940
Above 0 but less than 10,000	31,554	57.60%	2,317	5,590	31,554	2,130	1,364	14,235
More than 10,000	86	0.16%	6	33	86	22	4	44
Total	54,785	100%	2,584	6,349	54,762	2,606	1,561	17,219

Figure 1.3: Frequency table for Churn

churn_t1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	53414	97.5	97.5	97.5
	1.00	1371	2.5	2.5	100.0
	Total	54785	100.0	100.0	

Note: *0.00 means the customer is with the bank, and 1.00 is the customer who has left the bank.*

Figure 1.4: Frequency table for non-metric variables

Note: *For all product categories, 0 is referred to as not owning the financial product, and 1 is referred to as owning the financial product.*

Checkings Acc

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	23	.0	.0	.0
	1	54762	100.0	100.0	100.0
	Total	54785	100.0	100.0	

Credit card

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	48436	88.4	88.4	88.4
	1	6349	11.6	11.6	100.0
	Total	54785	100.0	100.0	

Loan

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	52201	95.3	95.3	95.3
	1	2584	4.7	4.7	100.0
	Total	54785	100.0	100.0	

Mortgage

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	53224	97.2	97.2	97.2
	1	1561	2.8	2.8	100.0
	Total	54785	100.0	100.0	

Inv Trust

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	52179	95.2	95.2	95.2
	1	2606	4.8	4.8	100.0
	Total	54785	100.0	100.0	

Savings

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	37566	68.6	68.6	68.6
	1	17219	31.4	31.4	100.0
	Total	54785	100.0	100.0	

Figure 1.5: Pearson Correlation test

Correlations

		Income	Number transactions	Age
Income	Pearson Correlation	1	.453**	-.005
	Sig. (2-tailed)		<.001	.356
	N	31640	31640	31638
Number transactions	Pearson Correlation	.453**	1	.085**
	Sig. (2-tailed)	<.001		<.001
	N	31640	54785	54783
Age	Pearson Correlation	-.005	.085**	1
	Sig. (2-tailed)	.356	<.001	
	N	31638	54783	54783

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 1.6: Final cluster – used for strategy recommendation

Legend	
0	Do not Own
1	Own

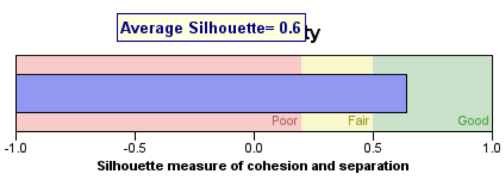
Cluster Number	Cluster Name	Size	Credit Card	Income	Inv Trust	Loan	Mortgage	Savings Account
1	Low Income - No Product Owners	32,263 (59.0%)	(0) 100%	357.46	(0) 100%	(0) 100%	(0) 100%	(0) 100%
5	Moderate Income - Savers	11,514 (21.0%)	(0) 100%	937.85	(0) 100%	(0) 100%	(0) 100%	(1) 100%
2	High Income - Credit Heavy Savers	4,687 (8.6%)	(1) 95.1%	1,673.67	(0) 100%	(0) 100%	(0) 100%	(0) 50.3%
3	High Income - Investors	2,583 (4.7%)	(0) 70.7%	1,524.51	(1) 100%	(0) 95.8%	(0) 93.0%	(1) 67.5%
6	High Income - Debtors	2,276 (4.2%)	(0) 66.4%	1,469.52	(0) 100%	(1) 100%	(0) 100%	(0) 59.8%
4	High Income - Home Owners	1,376 (2.5%)	(0) 75.6%	1,457.90	(0) 100%	(0) 86.0%	(1) 100%	(0) 51.2%

Figure 1.7: Two-step cluster analysis

With - Number of Transactions

Model Summary

Algorithm	TwoStep
Inputs	7
Clusters	6



Clusters

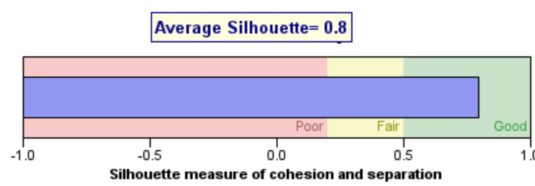
Input (Predictor) Importance: 1.0 0.8 0.6 0.4 0.2 0.0

Cluster	1	6	4	5	2	3
Label						
Description						
Size	46.2% (25258)	21.1% (11515)	13.1% (7162)	8.1% (4447)	6.7% (3659)	4.9% (2658)
Inputs	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 1 (100.0%)	Credit card 96 0 (69.9%)	Credit card 96 0 (71.1%)
	Income 96 129.74	Income 96 938.93	Income 96 1,244.72	Income 96 1,483.89	Income 96 1,464.09	Income 96 1,694.80
	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 1 (97.2%)
	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 1 (67.4%)	Loan96 0 (95.8%)
	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (62.5%)	Mortgage 96 0 (93.1%)
	Number transactions 96 51.77	Number transactions 96 216.84	Number transactions 96 272.75	Number transactions 96 307.87	Number transactions 96 313.57	Number transactions 96 279.03
	Savings 96 0 (100.0%)	Savings 96 1 (100.0%)	Savings 96 0 (100.0%)	Savings 96 1 (51.2%)	Savings 96 0 (56.5%)	Savings 96 1 (67.5%)

Without - Number of Transactions

Model Summary

Algorithm	TwoStep
Inputs	6
Clusters	6



Clusters

Input (Predictor) Importance: 1.0 0.8 0.6 0.4 0.2 0.0

Cluster	1	5	2	3	6	4
Label						
Description						
Size	59.0% (32263)	21.0% (11514)	8.6% (4687)	4.7% (2583)	4.2% (2276)	2.5% (1376)
Inputs	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 1 (95.1%)	Credit card 96 0 (70.7%)	Credit card 96 0 (66.4%)	Credit card 96 0 (75.6%)
	Income 96 357.46	Income 96 937.85	Income 96 1,673.67	Income 96 1,524.51	Income 96 1,469.52	Income 96 1,457.90
	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 1 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)
	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (95.8%)	Loan96 1 (100.0%)	Loan96 0 (86.0%)
	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (93.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 1 (100.0%)
	Savings 96 0 (100.0%)	Savings 96 1 (100.0%)	Savings 96 0 (50.3%)	Savings 96 1 (67.5%)	Savings 96 0 (59.8%)	Savings 96 0 (51.2%)

Figure 1.8: ANOVA test

Descriptives

Income 96

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
1	32263	357.46	567.805	3.161	351.26	363.65	0	3265
2	4687	1673.67	1458.609	21.305	1631.91	1715.44	0	9908
3	2583	1524.51	1304.987	25.677	1474.16	1574.86	0	9631
4	1376	1457.90	1141.917	30.784	1397.51	1518.29	0	9581
5	11514	937.85	786.857	7.333	923.48	952.23	0	4693
6	2276	1469.52	1020.223	21.385	1427.58	1511.46	0	7333
Total	54699	721.48	944.612	4.039	713.56	729.39	0	9908

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Income 96	Based on Mean	1495.398	5	54693	<.001
	Based on Median	1346.057	5	54693	<.001
	Based on Median and with adjusted df	1346.057	5	39109.880	<.001
	Based on trimmed mean	1421.119	5	54693	<.001

ANOVA

Income 96

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12749334227	5	2549866845.4	3867.736	<.001
Within Groups	36057237339	54693	659266.037		
Total	48806571566	54698			

Robust Tests of Equality of Means

Income 96

	Statistic[a]	df1	df2	Sig.
Welch	2646.589	5	6525.668	<.001

a. Asymptotically F distributed.

Multiple Comparisons

Dependent Variable: Income 96

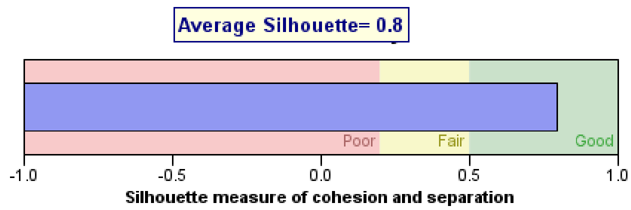
	(I) TwoStep Cluster Number	(J) TwoStep Cluster Number	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tamhane	1	2	-1316.218*	21.539	<.001	-1379.31	-1253.13
		3	-1167.049*	25.871	<.001	-1242.86	-1091.24
		4	-1100.441*	30.946	<.001	-1191.20	-1009.68
		5	-580.395*	7.985	<.001	-603.78	-557.01
		6	-1112.064*	21.617	<.001	-1175.42	-1048.71
	2	1	1316.218*	21.539	<.001	1253.13	1379.31
		3	149.168*	33.365	<.001	51.44	246.89
		4	215.777*	37.438	<.001	106.07	325.48
		5	735.822*	22.532	<.001	669.83	801.82
		6	204.154*	30.187	<.001	115.74	292.57
	3	1	1167.049*	25.871	<.001	1091.24	1242.86
		2	-149.168*	33.365	<.001	-246.89	-51.44
		4	66.609	40.087	.782	-50.85	184.06
		5	586.654*	26.704	<.001	508.41	664.90
		6	54.985	33.416	.794	-42.90	152.87
	4	1	1100.441*	30.946	<.001	1009.68	1191.20
		2	-215.777*	37.438	<.001	-325.48	-106.07
		3	-66.609	40.087	.782	-184.06	50.85
		5	520.045*	31.645	<.001	427.25	612.84
		6	-11.624	37.483	1.000	-121.47	98.22
	5	1	580.395*	7.985	<.001	557.01	603.78
		2	-735.822*	22.532	<.001	-801.82	-669.83
		3	-586.654*	26.704	<.001	-664.90	-508.41
		4	-520.045*	31.645	<.001	-612.84	-427.25
		6	-531.669*	22.607	<.001	-597.91	-465.42
	6	1	1112.064*	21.617	<.001	1048.71	1175.42
		2	-204.154*	30.187	<.001	-292.57	-115.74
		3	-54.985	33.416	.794	-152.87	42.90
		4	11.624	37.483	1.000	-98.22	121.47
		5	531.669*	22.607	<.001	465.42	597.91

*. The mean difference is significant at the 0.05 level.

Figure 1.9: SPSS output for the final cluster also used for recommendations.

Model Summary

Algorithm	TwoStep
Inputs	6
Clusters	6



Clusters

Input (Predictor) Importance
1.0 0.8 0.6 0.4 0.2 0.0

Cluster	1	5	2	3	6	4
Label						
Description						
Size	59.0% (32263)	21.0% (11514)	8.6% (4687)	4.7% (2583)	4.2% (2276)	2.5% (1376)
Inputs	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 1 (95.1%)	Credit card 96 0 (70.7%)	Credit card 96 0 (66.4%)	Credit card 96 0 (75.6%)
	Income 96 357.46	Income 96 937.85	Income 96 1,673.67	Income 96 1,524.51	Income 96 1,469.52	Income 96 1,457.90
	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 1 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)
	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (95.8%)	Loan96 1 (100.0%)	Loan96 0 (86.0%)
	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (93.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 1 (100.0%)
	Savings 96 0 (100.0%)	Savings 96 1 (100.0%)	Savings 96 0 (50.3%)	Savings 96 1 (67.5%)	Savings 96 0 (59.8%)	Savings 96 0 (51.2%)

Figure 2.0: Chi-square test

Loan96 * TwoStep Cluster Number

Crosstab

Count

		TwoStep Cluster Number						
		1	2	3	4	5	6	Total
Loan96	0	32263	4686	2474	1184	11514	0	52121
	1	0	1	109	192	0	2276	2578
Total		32263	4687	2583	1376	11514	2276	54699

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	48673.321[a]	5	<.001
Likelihood Ratio	18748.833	5	<.001
Linear-by-Linear Association	9857.732	1	<.001
N of Valid Cases	54699		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 64.85.

Inv Trust 96 * TwoStep Cluster Number

Crosstab

Count

		TwoStep Cluster Number						
		1	2	3	4	5	6	Total
Inv Trust 96	0	32263	4686	0	1376	11514	2276	52115
	1	0	1	2583	0	0	0	2584
Total		32263	4687	2583	1376	11514	2276	54699

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	54676.787[a]	5	<.001
Likelihood Ratio	20800.412	5	<.001
Linear-by-Linear Association	407.598	1	<.001
N of Valid Cases	54699		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 65.00.

Savings 96 * TwoStep Cluster Number

Crosstab

Count

		TwoStep Cluster Number						
		1	2	3	4	5	6	Total
Savings 96	0	32263	2357	839	705	0	1360	37524
	1	0	2330	1744	671	11514	916	17175
Total		32263	4687	2583	1376	11514	2276	54699

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	42492.324[a]	5	<.001
Likelihood Ratio	53344.632	5	<.001
Linear-by-Linear Association	34406.561	1	<.001
N of Valid Cases	54699		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 432.05.

Credit card 96 * TwoStep Cluster Number

Crosstab

Count

		TwoStep Cluster Number						
		1	2	3	4	5	6	Total
Credit card 96	0	32263	228	1826	1040	11514	1512	48383
	1	0	4459	757	336	0	764	6316
Total		32263	4687	2583	1376	11514	2276	54699

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	39879.927[a]	5	<.001
Likelihood Ratio	29759.765	5	<.001
Linear-by-Linear Association	364.522	1	<.001
N of Valid Cases	54699		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 158.88.

Mortgage 96 * TwoStep Cluster Number

Crosstab

Count

		TwoStep Cluster Number						
		1	2	3	4	5	6	Total
Mortgage 96	0	32263	4687	2402	0	11514	2276	53142
	1	0	0	181	1376	0	0	1557
Total		32263	4687	2583	1376	11514	2276	54699

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	48612.615[a]	5	<.001
Likelihood Ratio	12840.961	5	<.001
Linear-by-Linear Association	1245.490	1	<.001
N of Valid Cases	54699		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 39.17.

Figure 2.1: Test clusters

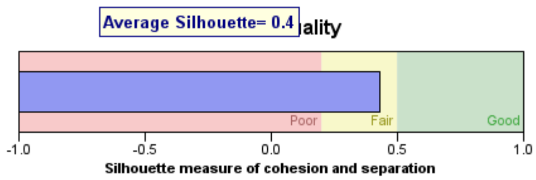
<u>*Test Cluster 1:*</u>

Variables:

- *Age*
- *Credit Card*
- *Income*
- *Investment Trust*
- *Loans*
- *Mortgage*
- *Number of Transactions*
- *Savings Account*

Model Summary

Algorithm	TwoStep
Inputs	8
Clusters	6



Clusters

Input (Predictor) Importance
1.0 0.8 0.6 0.4 0.2 0.0

Cluster	1	2	4	5	6	3
Label						
Description						
Size	31.6% (16331)	26.8% (13831)	20.4% (10556)	8.6% (4425)	8.3% (4272)	4.3% (2235)
Inputs	Age 96 29.30	Age 96 53.70	Age 96 46.96	Age 96 40.09	Age 96 43.03	Age 96 49.04
	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 1 (100.0%)	Credit card 96 0 (71.1%)	Credit card 96 0 (71.7%)
	Income 96 48.89	Income 96 697.98	Income 96 936.85	Income 96 1,483.41	Income 96 1,745.38	Income 96 1,479.74
	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (93.6%)	Inv Trust 96 1 (100.0%)
	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 1 (60.3%)	Loan96 0 (100.0%)
	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (63.8%)	Mortgage 96 0 (100.0%)
	Number transactions 96 36.23	Number transactions 96 171.34	Number transactions 96 220.35	Number transactions 96 307.89	Number transactions 96 331.75	Number transactions 96 263.24
	Savings 96 0 (100.0%)	Savings 96 0 (100.0%)	Savings 96 1 (100.0%)	Savings 96 1 (51.1%)	Savings 96 0 (56.6%)	Savings 96 1 (67.1%)

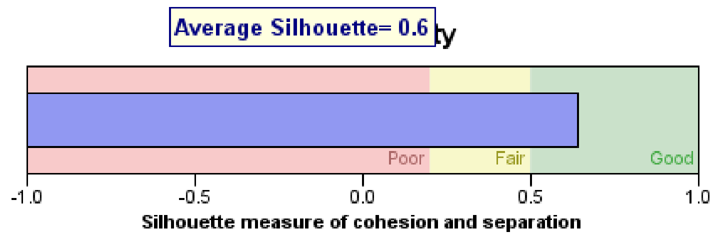
Test cluster 2:

Variables:

- *Credit Card*
- *Income*
- *Investment Trust*
- *Loans*
- *Mortgage*
- *Number of Transactions*
- *Savings Account*

Model Summary

Algorithm	TwoStep
Inputs	7
Clusters	6



Clusters

Input (Predictor) Importance

1.0 0.8 0.6 0.4 0.2 0.0

Cluster	1	6	4	5	2	3
Label						
Description						
Size	46.2% (25258)	21.1% (11515)	13.1% (7162)	8.1% (4447)	6.7% (3659)	4.9% (2658)
Inputs	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 1 (100.0%)	Credit card 96 0 (69.9%)	Credit card 96 0 (71.1%)
	Income 96 129.74	Income 96 938.93	Income 96 1,244.72	Income 96 1,483.89	Income 96 1,464.09	Income 96 1,694.80
	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 1 (97.2%)
	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 1 (67.4%)	Loan96 0 (95.8%)
	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (62.5%)	Mortgage 96 0 (93.1%)
	Number transactions 96 51.77	Number transactions 96 216.84	Number transactions 96 272.75	Number transactions 96 307.87	Number transactions 96 313.57	Number transactions 96 279.03
	Savings 96 0 (100.0%)	Savings 96 1 (100.0%)	Savings 96 0 (100.0%)	Savings 96 1 (51.2%)	Savings 96 0 (56.5%)	Savings 96 1 (67.5%)

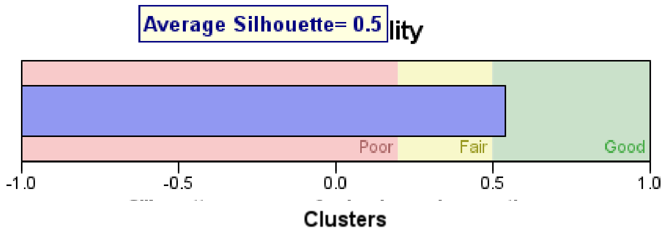
Test cluster 3:

Variables:

- *Age*
- *Credit Card*
- *Income*
- *Investment Trust*
- *Loans*
- *Mortgage*
- *Savings Account*

Model Summary

Algorithm	TwoStep
Inputs	7
Clusters	6



Input (Predictor) Importance

1.0 0.8 0.6 0.4 0.2 0.0

Cluster	3	1	4	6	2	5
Label						
Description						
Size	31.9% (16477)	26.9% (13876)	20.5% (10595)	8.5% (4416)	7.1% (3671)	5.1% (2615)
Inputs	Age 96 53.87	Age 96 24.64	Age 96 46.96	Age 96 40.08	Age 96 42.54	Age 96 48.77
	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 1 (100.0%)	Credit card 96 0 (70.1%)	Credit card 96 0 (70.3%)
	Income 96 580.47	Income 96 107.63	Income 96 948.26	Income 96 1,471.04	Income 96 1,485.85	Income 96 1,764.40
	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 1 (95.9%)
	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 1 (67.1%)	Loan96 0 (95.7%)
	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (62.9%)	Mortgage 96 0 (93.0%)
	Savings 96 0 (100.0%)	Savings 96 0 (100.0%)	Savings 96 1 (100.0%)	Savings 96 1 (51.1%)	Savings 96 0 (57.0%)	Savings 96 1 (66.6%)

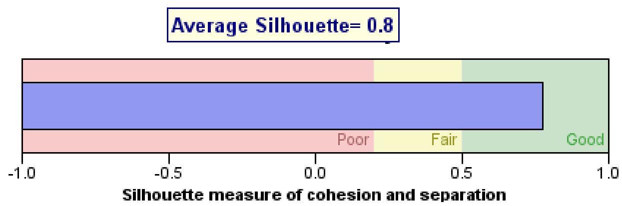
Figure 2.2: Test clusters

Test cluster 1:

- *Six input variables, five forced clusters*

Model Summary

Algorithm	TwoStep
Inputs	6
Clusters	5



Clusters

Input (Predictor) Importance
1.0 0.8 0.6 0.4 0.2 0.0

Cluster	1	5	2	4	3
Label					
Description					
Size	58.9% (32234)	21.0% (11514)	8.5% (4628)	6.8% (3740)	4.7% (2583)
Inputs	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 1 (96.3%)	Credit card 96 0 (70.6%)	Credit card 96 0 (70.7%)
	Income 96 354.88	Income 96 937.85	Income 96 1,651.16	Income 96 1,509.88	Income 96 1,524.51
	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 1 (100.0%)
	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 1 (66.0%)	Loan96 0 (95.8%)
	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (63.2%)	Mortgage 96 0 (93.0%)
	Savings 96 0 (100.0%)	Savings 96 1 (100.0%)	Savings 96 1 (50.3%)	Savings 96 0 (57.6%)	Savings 96 1 (67.5%)

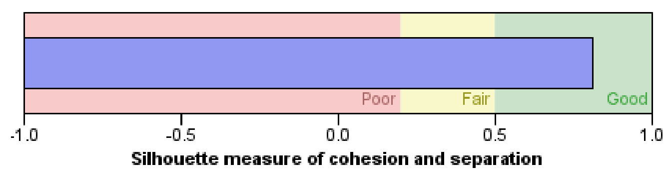
Test cluster 2:

- *Six inputs, seven forced clusters*

Model Summary

Algorithm	TwoStep
Inputs	6
Clusters	7

Cluster Quality



Clusters

Input (Predictor) Importance

1.0 0.8 0.6 0.4 0.2 0.0

Cluster	1	5	2	3	7	6	4
Label							
Description							
Size	58.6% (32063)	21.1% (11532)	4.7% (2595)	4.7% (2582)	4.2% (2277)	4.2% (2274)	2.5% (1376)
Inputs	Credit card 96 0 (100.0%)	Credit card 96 0 (100.0%)	Credit card 96 1 (84.2%)	Credit card 96 0 (70.7%)	Credit card 96 0 (66.4%)	Credit card 96 1 (100.0%)	Credit card 96 0 (75.6%)
	Income 96 341.43	Income 96 944.56	Income 96 1,781.44	Income 96 1,521.88	Income 96 1,472.67	Income 96 1,632.61	Income 96 1,457.90
	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (99.9%)	Inv Trust 96 1 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)	Inv Trust 96 0 (100.0%)
	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (95.8%)	Loan96 1 (100.0%)	Loan96 0 (100.0%)	Loan96 0 (86.0%)
	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (93.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 0 (100.0%)	Mortgage 96 1 (100.0%)
	Savings 96 0 (100.0%)	Savings 96 1 (100.0%)	Savings 96 0 (98.6%)	Savings 96 1 (67.5%)	Savings 96 0 (59.7%)	Savings 96 1 (100.0%)	Savings 96 0 (51.2%)

References

Stats NZ. (2023). Life Expectancy. Stat NZ https://www.stats.govt.nz/topics/life-expectancy

OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. https://chat.openai.com/chat

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. Journal of marketing, 80(6), 69-96.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate Data Analysis (8th ed.). Cengage Learning.

Shih, M. Y., Jheng, J. W., & Lai, L. F. (2010). A two-step method for clustering mixed categorical and numeric data. Journal of Applied Science and Engineering, 13(1), 11-19.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. Pattern Recognition, 46(1), 243-256.

Dolnicar, S., & Grün, B. (2008). Challenging "factor–cluster segmentation". Journal of Travel Research, 47(1), 63-71.

Demirguc-Kunt, A., Klapper, L., Singer, D. and Ansar, S., 2018. The Global Findex Database 2017: Measuring financial inclusion and the fintech revolution. World Bank Publications.

Keiningham, T. L., Cooil, B., Aksoy, L., Andreassen, T. W., & Weiner, J. (2007). The value of different customer satisfaction and loyalty metrics in predicting customer retention, recommendation, and share-of-wallet. Managing service quality: An international Journal, 17(4), 361-384.

Kim, Jung-Hwan, and Minjeong Kim. "Conceptualization and assessment of E-service quality for luxury brands." The Service Industries Journal 40, no. 5-6 (2020): 436-470.

Aksoy, L., Buoye, A., Aksoy, P., & Larivière, B. (2008). A Cross-National Investigation of the Satisfaction and Loyalty Linkage for Mobile Telephony Services across Eight Countries. Journal of Interactive Marketing, 22(3), 5-24.

AI Acknowledgement

I would like to acknowledge the use of artificial intelligence, specifically OpenAI's ChatGPT, in assisting with the writing and research of this report. Parts of the assignment, like describing cluster profiles, and providing pointers for potential recommendations, were supported by AI. The AI provided valuable insights and suggestions, which were instrumental in shaping the content and ensuring a comprehensive analysis of the cluster evaluation.