TE KUNENGA KI PŪREHUROA
MASSEY UNIVERSITY
UNIVERSITY OF NEW ZEALAND

TE KURA WHAI PAKIHI
MASSEY BUSINESS SCHOOL

125788

Big Data in Finance and Banking

Customer Segmentation and Probit Regression Analysis of Financial Behaviour: Insights from Visual Analytics

Luis Vieira
23012096

September, 2024

Index

1. Introduction

This analysis explores the relationship between gambling behaviour, financial capability or literacy, and financial hardship. Financial hardship and distress are complex issues, often worsened by gambling behaviour. Marko *et al.* (2023) explain that gamblers and those affected by gambling often do not recognise the financial risks until substantial harm occurs, leading to debt and long-term economic hardship. Similarly, Armstrong and Thomas (2017) point out to changes in gambling behaviour over time, with a considerable increase in spending on high risk gambling like electronic gaming machines (EGM). Even though there is a decrease in overall gambling activities, increased spending on EGM suggests a growing risk of financial harm.
On the other hand, financial capability plays an important role in family financial outcomes. Worthington (2006) says financial capability is higher among older, educated, and wealthier groups, whereas those with lower education levels, non-English speakers and women tend to display lower financial literacy. Bourova et al. (2018), however, caution that financial literacy alone is not enough to ease financial hardship, as overconfidence in financial capability can increase financial difficulties, especially for those with lower education levels.
Koomson et al. (2022) explore the gender differences in gambling related financial stress, showing that men often face higher gambling severity and harm, while women show higher levels of financial stress. Concluding that promoting financial resilience, can reduce the impact of gambling on financial hardship.
Given the difficulty in understanding the impacts and relationships between financial capability, hardship, and the effects of gambling behaviours, a different approach is needed, such as the use of segmentation analysis. Scroggins et al. (2006) suggests a needs-based segmentation to better understand financial behaviour, which can help institutions predict credit risk and customise products. The segmentation can help us identify high-risk and vulnerable groups of individuals, facing financial struggles due to debt and gambling harm.
In summary, the relationships between financial hardship, financial literacy, and gambling behaviour are complex. Segmentation, by providing a more nuanced understanding of customer profiles, plays a crucial role in identifying and addressing these challenges, providing a deeper understanding of consumer financial behaviour, financial institutions can better predict risks and offer more targeted support to those most in need.

2. Data, Methodology, and Descriptive Statistics

2.1 Dataset Description

This dataset captures four key dynamics: financial capability, financial hardship, gambling behaviour, and demographics. It consists of categorical, binary, and continuous variables that measure various aspects of individuals' financial behaviour and personal circumstances.

Table 1: Financial Capability Variables

Variable Name	Type	Description	Range/Values
fcdealbank	Categorical	Comfort level with dealing with banks and financial institutions	0-6
fcbalspend	Categorical	Ability to balance spending and savings	0-6
fcconffd	Categorical	Confidence in financial decisions	0-6
fcgooddfm	Categorical	Capability of managing day-to-day financial matters	0-6
fcemergsav	Categorical	Preparedness for emergencies by saving	0-6
fcfingoal	Categorical	Long-term financial goal-setting	0-6
fcorganmm	Categorical	Organisation in managing daily finances	0-6
fcfincontract	Categorical	Understanding of financial contracts	0-6
fcfinwatch	Categorical	Close watch on personal financial affairs	0-6
fincapscore	Continuous	Financial capability average score (of categorical variables)	0-6
fincapscoresum	Continuous	Financial capability score (sum of financial capability questions)	0-54
leastfincap	Binary	Least financial capability group	0 = No, 1 = Yes

Table 2: Financial Hardship Variables

Variable Name	Type	Description	Range/Values
finbillnpaid	Binary	Could not pay utility bills on time	0 = No, 1 = Yes
finaskhelp	Binary	Asked for financial help from friends or family	0 = No, 1 = Yes
finrentmornpaid	Binary	Could not pay mortgage on time	0 = No, 1 = Yes
finsoldpawn	Binary	Pawned or sold something	0 = No, 1 = Yes
finnoheat	Binary	Unable to heat home	0 = No, 1 = Yes
finwwithoutfood	Binary	Went without meals	0 = No, 1 = Yes
finwelfarehelp	Binary	Sought help from welfare	0 = No, 1 = Yes
fhardship	Continuous	Financial hardship score (sum of hardship-related questions)	0-7

Table 3: Gambling Behaviour Variables

Variable Name	Type	Description	Range/Values
gamborrowed	Categorical	Borrowed money or sold items to gamble	0-3
gamcriticize	Categorical	Received criticism for gambling	0-3
gamfindif	Categorical	Gambling caused financial issues	0-3
gamguilt	Categorical	Felt guilty about gambling	0-3
gamhealthprob	Categorical	Gambling caused health problems	0-3
gamnotafford	Categorical	Bet more than could afford to lose	0-3
gaharm	Continuous	Gambling harm score (sum of gambling harm-related questions)	0-18

Table 4: Demographic Variables

Variable Name	Type	Description	Range/Values
age	Continuous	Age of the respondent	15-99 years
employed	Binary	Employment status	0 = No, 1 = Yes
incomepos	Continuous	Total income in dollars	3-2,022,000
married	Binary	Marital status	0 = No, 1 = Yes
highestedu	Categorical	Highest education level achieved	0-6 (where 0 = less than Year 12; 1 = Year 12; 2 = Cert III or IV; 3 = Advanced diploma, diploma ; 4 = Bachelor honours ; 5 = Graduate diploma/certificate; 6 = Postgraduate degree)
ownhome	Binary	Home ownership	0 = No, 1 = Yes
segment	Categorical	Segment cluster identifier after clustering	1-...

2.2 Methodology

For the segmentation analysis, even though both SAS Miner and SAS Viya were utilised to explore and segment the data, I decided to leverage SAS Viya Model Studio to create Segmentation pipelines similar to what SAS Miner does, while visual analysis and regression analysis was conducted only in SAS Viya.
Given that the dataset had less than 5% missing values, no imputation was performed to avoid introducing potential distortions. Similarly, no outliers were removed to maintain the natural integrity of the dataset. To prepare the data for segmentation, I standardised the ordinal variables (using z-scores) whilst leaving the binary variables as is, to avoid risking distortions, and to ensure that all features had equal weight in the clustering process. Additionally, I applied a logarithmic transformation to the *incomepos* variable to address skewness/kurtosis and achieve a more normal distribution. I also made the decision to separate and exclude individuals aged over 65, as retirees tend to display different financial behaviours and dynamics, which would have skewed the segmentation results.
For the segmentation itself, SAS Miner uses Ward method (hierarchical clustering), whilst SAS Viya's uses K-means as clustering algorithm, with Euclidean distance as the distance metric, allowing to compare and assess robustness of the segmentations. This method allowed for the identification of distinct customer segments based on their financial capability, hardship, and gambling behaviours, before running logistic regressions (Probit) by segment in SAS Viya to evaluate and interpret direction and strength of relationship of variables.

2.3 Descriptive Statistics

Table 5: Descriptive Statistics

Variable	Mean	Std Dev	N	Missing	Min	Median	Max	Skewness	Kurtosis
age	46.07	18.83	15235	0	15	45	**99**	0.24	-0.90
employed	0.63	0.48	15235	0	0	1	1	-0.54	-1.71
fcbalspend	3.89	1.62	15218	17	0	4	6	-0.54	-0.45
fcconffd	4.11	1.39	15213	22	0	4	6	-0.67	0.19
fcdealbank	3.90	1.72	15235	0	0	4	6	-0.61	-0.54
fcemergsav	4.01	1.78	15216	19	0	4	6	-0.67	-0.58
fcfincontract	4.65	1.39	15191	44	0	5	6	-1.16	1.08
fcfingoal	3.61	1.67	15198	37	0	4	6	-0.36	-0.68
fcfinwatch	4.20	1.53	15221	14	0	5	6	-0.74	-0.11
fcgooddfm	4.15	1.53	15213	22	0	4	6	-0.72	-0.08
fcorganmm	3.83	1.62	15217	18	0	4	6	-0.45	-0.59
fhardship	0.43	1.07	14957	**278**	0	**0**	7	3.11	10.66
finaskhelp	0.11	0.32	15060	175	0	0	1	2.42	3.85
finbillnpaid	0.11	0.32	15063	172	0	0	1	2.43	3.89
fincapscore	4.04	1.24	15235	0	0	**4.11**	6	-0.53	-0.03
fincapscoresum	36.39	11.12	15076	159	0	37	54	-0.53	-0.03
finnoheat	0.03	0.16	15049	186	0	0	1	5.80	31.65
finrentmornpaid	0.05	0.23	15039	196	0	0	1	3.96	13.67
finsoldpawn	0.05	0.23	15047	188	0	0	1	3.91	13.30
finwelfarehelp	0.03	0.18	15051	184	0	0	1	5.07	23.66
finwwithoutfood	0.04	0.19	15057	178	0	0	1	4.89	21.89
gaharm	0.17	0.89	15235	0	0	0	18	9.16	**112.78**
gamborrowed	0.01	0.12	15235	0	0	0	3	14.94	272.45
gamcriticize	0.03	0.17	15235	0	0	0	3	9.27	107.01
gamfindif	0.02	0.15	15235	0	0	0	3	11.68	169.53
gamguilt	0.05	0.25	15235	0	0	0	3	6.77	56.25
gamhealthprob	0.02	0.17	15235	0	0	0	3	10.67	140.96
gamnotafford	0.05	0.22	15235	0	0	0	3	6.39	52.06
highestedu	2.16	1.82	15228	7	0	2	6	0.48	-0.80
incomepos	56036.67	64061.69	14669	**566**	3	**42000**	**2022000**	**8.79**	**172.90**
leastfincap	0.90	0.30	15235	0	0	1	1	-2.65	5.01
married	0.48	0.50	15234	1	0	0	1	0.06	-2.00
ownhome	0.69	0.46	15220	15	0	1	1	-0.81	-1.35

The overview of the descriptive statistics Table 1, shows the median age is 45, and although there is a significant group of retirees aged over 65, up to 99 years, I believe this group shows different financial dynamics. Thus, in this segmentation analysis, I will primarily focus on working age individuals, aged 65 or younger.

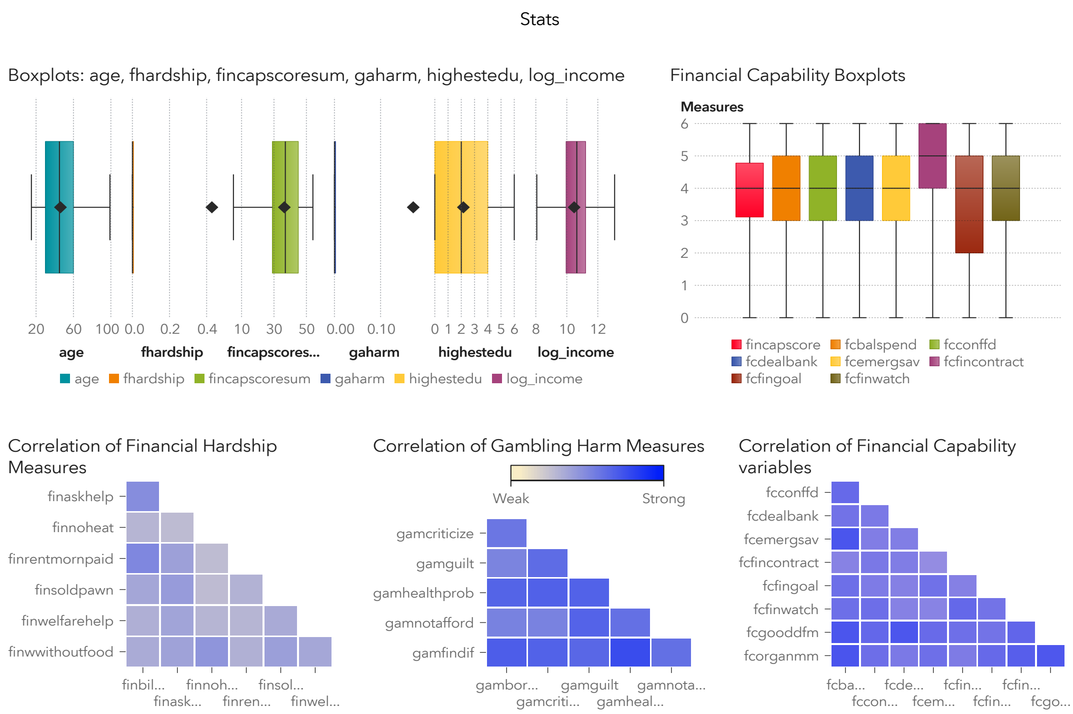
The income variable is right-skewed and has high kurtosis (more extreme values), median of $42,000 and maximum of $202,200, suggesting that a logarithmic transformation could provide a better fit for analysis.

In terms of educational background, the median level places individuals between a Certificate III/IV and an Advanced Diploma, with relatively few having qualifications higher than a Bachelor's degree.

The data on financial capability and literacy shows relatively strong results, with median values above 4 out of 6 across all financial capability variables. Individuals tend to be most literate when it comes to understanding financial contracts, but they demonstrate weaker performance in setting long-term financial goals (Image 1 Financial Capability Boxplots).

Financial hardship is present, luckily not widespread, with 278 missing values (less than 2%). The mean financial hardship score of 0.43, coupled with the skewness and lower values in related variables (all below 0.11), suggests that around 40% of individuals experience some form of financial distress. However, a smaller proportion faces severe financial hardship. Note that those struggling with utility bill payments are correlated at 49.33% with also not paying rent or mortgage on time (extracted from SAS Viya Image 1 Correlation of Financial Hardship Measures).

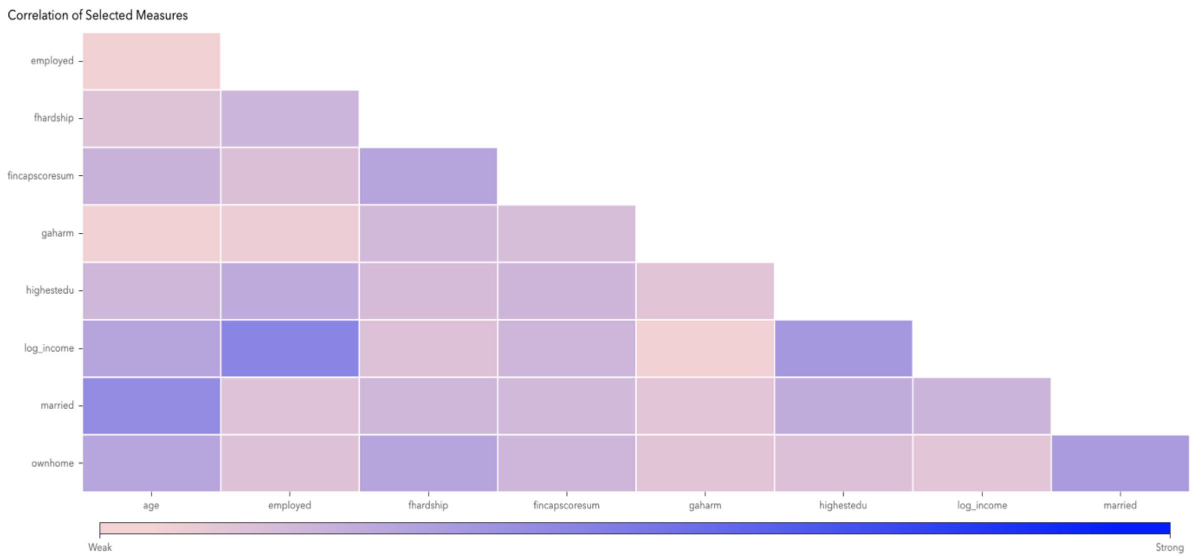
Image 1: Boxplots and Correlation Matrices of Main Variables



Regarding gambling harm, approximately 17% of individuals exhibit some form of gambling-related problematic behaviour, though they are largely outliers. Among these individuals, gambling related financial difficulties are strongly correlated with health problems (77.71%) and borrowing money for gambling (69.6%). Furthermore, individuals unable to afford food and heating show a 42.68% correlation (extracted from SAS Viya Image 1 Correlation of Gambling Harm Measures).

The financial hardship score variable also shows negative correlations of about -25.6%, although weak, with financial capability score sum and owning a home (extracted from SAS Viya Image 2 Correlation Table).

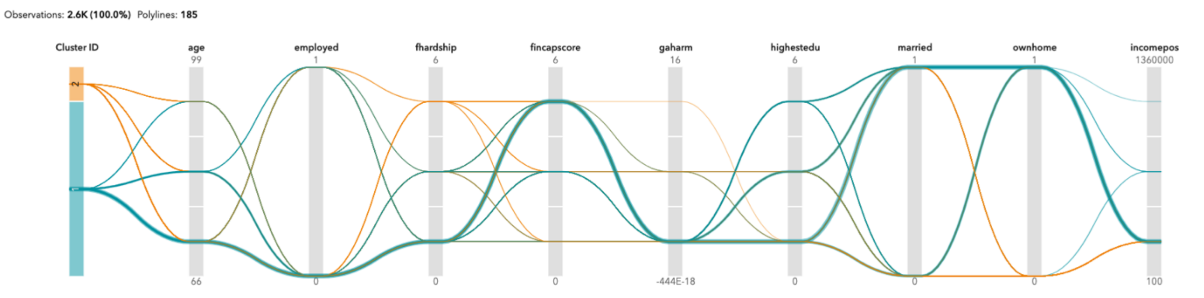
Image 2: Correlation Table of Selected Variables



3. Segmentation Analysis

As previously mentioned, my analysis mainly focuses on working age individuals, as literature illustrates these retirees display a different financial and behaviour dynamics.
Prior to proceeding to the working age segmentation, a brief analysis (Image 3) shows that retirees (over 65) can likely be divided into 2 segments. On one hand those financially secured (cluster 1, 2169 observations) enjoying their retirement with little to no financial hardship, likely higher educated and financially capable, married, income close to the mean, and most own their home.
On the other hand, a minority group (431 observations) that might be more vulnerable to financial hardships, less likely to be married, mostly do not own their home, have less disposable income, and about a third gambling away their savings (cluster 2).

Image 3: Retirees Clustering (SAS Viya)



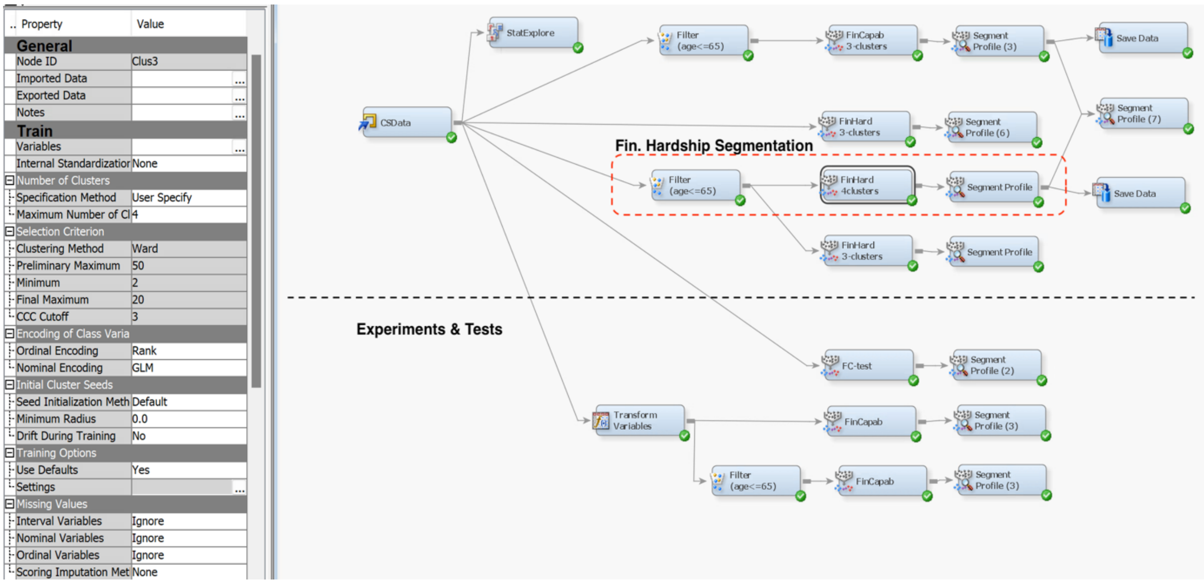
Centroids | Cluster Summary | Model Information | Within Cluster Statistics | Iteration History | Standardization | Interval information | Parallel Coordinates Plot

Cluster ID	age	employed	fhardship	fincapscore	gaharm	highestedu	married	ownhome	incomepos
1	74.09959	0.129553	0.07515	4.723067	0.098893	1.730751	0.648686	0.999078	39235.08
2	75.56613	0.083527	0.454756	4.49884	0.271462	1.211137	0.389791	0.039443	26131.05

3.1 Financial Hardship Segmentation Analysis

Now, moving towards our main analysis of the working age individuals. I initially ran the clustering and segmentation analysis in both SAS Miner and SAS Viya, and although results were similar, I decided to use Miner for the segmentation of financial hardship. Given that hierarchical clustering (Miner) is generally better than K-means (Viya) with Euclidean distances for binary data, as hierarchical methods can better handle the structure of binary variables.
Missing values were excluded and not imputed, also no outliers were removed, whereas non-binary variables were standardised.

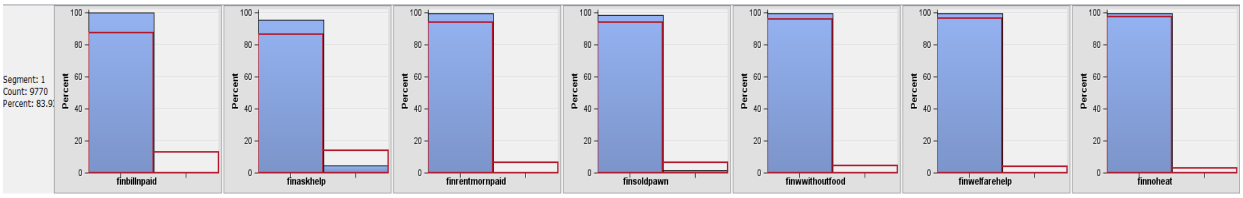
Image 4: SAS Miner SEMMA[1] Dashboard



Note. File name: LuisV_23012096_A1-Miner.xml

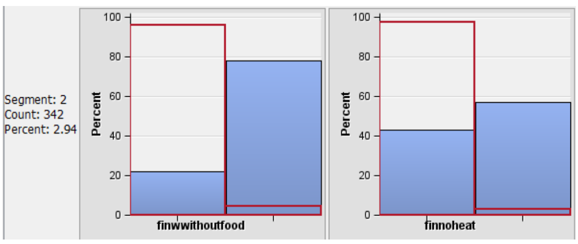
Having experimented with different number of clusters, I decided that 4 clusters/segments provided a more insightful segmentation than 3 clusters but without becoming too detailed like with higher number of clusters.

Segment 1: Financially Secure



The data clearly shows this segment faces minimal to no financial distress, so this appears to represent financially secure individuals. They tend to manage their finances well, with no trouble meeting financial obligations. Life is stable and comfortable for these individuals.

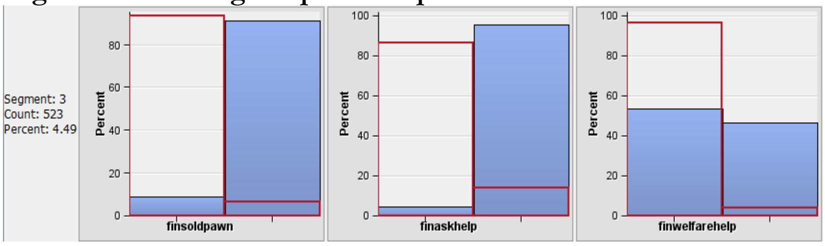
Segment 2: Highly Vulnerable



This segment seems to be facing extreme financial hardship, unable to afford basic necessities like food and heating. They are in a very precarious position, making them very vulnerable to financial and economic downturns.

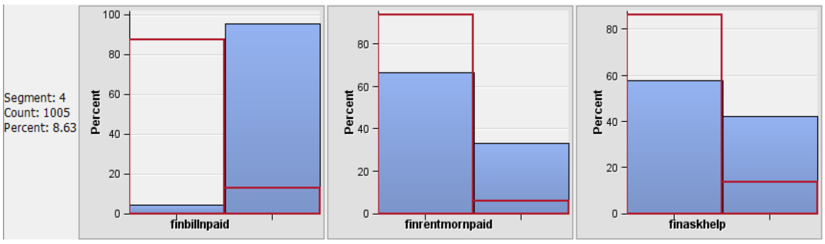
[1] SAS Miner SEMMA approach: Sample, Explore, Modify, Model, Assess

Segment 3: Seeking Help to Escape Distress



This segment is clearly financially distressed, but likely actively seeking ways to improve their situation by asking for help from family, friends, and welfare. They are taking significant measures, such as selling or pawning possessions, to regain financial stability.

Segment 4: Responding to Financial Shock



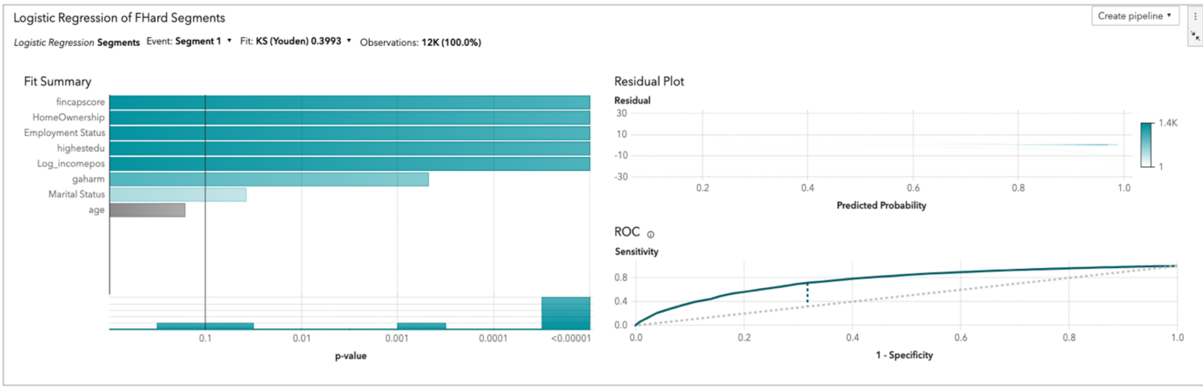
This segment likely faced a sudden financial shock, such as job loss, health issue, increased mortgage repayments and/or other unexpected expenses. They might not yet fully dependent on external assistance but may be managing temporary instability with the support of family or friends while trying to maintain financial control.

After describing the key characteristics of each segment, it is essential to further understand the underlying factors that contribute to financial hardship in these segments. To do so, I conducted a logistic regression (Probit) analysis for each segment. This allows to quantify how different variables (employment status, home ownership, marital status, financial capability, educational levels, income (log), gambling behaviour and age), impact the likelihood of belonging to each segment.

Nonetheless segment 1 are financially secured, I start presenting the results of this analysis with this segment, the most financially secured.

Note that this regression analysis was conducted in SAS Viya, but it seems that the Marginal Effects estimation has yet to be implemented, thus interpretation of direction and strength of the relationship is based on the coefficients and their values, whereas statistical significance is based on p-values less than 0.1.

Image 8: Segment 1 – Financially Secure – Logistic Regression (Probit)



Model Information | Dimensions | Response Profile | Iteration History | Convergence | Fit Statistics | Parameter Estimates | Type III Test | Confusion Matrix | Lift | ROC | Cutoff Statistics | Misclassification | Assessment Statistics

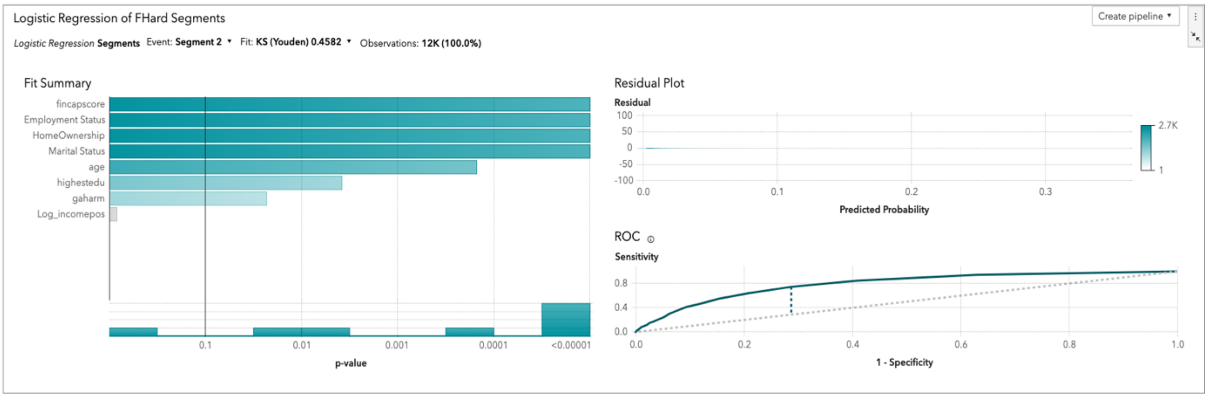
Parameter	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	0.458648	0.149317	9.435011	0.00213
Employment Status Employed	0.386431	0.038237	102.1339	<0.00001
Employment Status Not Employed	0	.	.	.
HomeOwnership No Home Own	-0.5342	0.032457	270.8791	<0.00001
HomeOwnership Own Home	0	.	.	.
Marital Status Married	0.071042	0.034196	4.315916	0.03776
Marital Status Not Married	0	.	.	.
age	0.001733	0.001242	1.947862	0.16282
gaharm	-0.04773	0.013663	12.20408	0.00048
fincapscore	0.27554	0.012682	472.066	<0.00001
highestedu	0.071883	0.009395	58.54709	<0.00001
Log_incomepos	-0.07488	0.015934	22.08267	<0.00001

Key Variables, Direction, Strength and Significance:

- Employment Status (Employed): Positive coefficient (0.386, highly significant)
- Homeownership (No Homeownership): Negative coefficient (-0.534, highly significant)
- Marital Status (Married): Positive coefficient (0.071, significant)
- Financial Capability (fincapscore): Positive coefficient (0.276, highly significant)
- Log of Income: Negative coefficient (-0.0749, highly significant)
- Age is not statistically significant

The estimates show that employment, home ownership, and marital status are strong factors in Segment 1. Financial capability has a positive effect, suggesting that individuals in this segment have high financial literacy and good money management skills, making them financially secure, as expected. Additionally, home ownership and being employed play an important role in maintaining financial stability, with individuals that do not have home ownership and lower income individuals being less likely to belong to this segment.

Image 5: Segment 2 – Highly Vulnerable – Logistic Regression (Probit)



Model Information | Dimensions | Response Profile | Iteration History | Convergence | Fit Statistics | Parameter Estimates | Type III Test | Confusion Matrix | Lift | ROC | Cutoff Statistics | Misclassification | Assessment Statistics

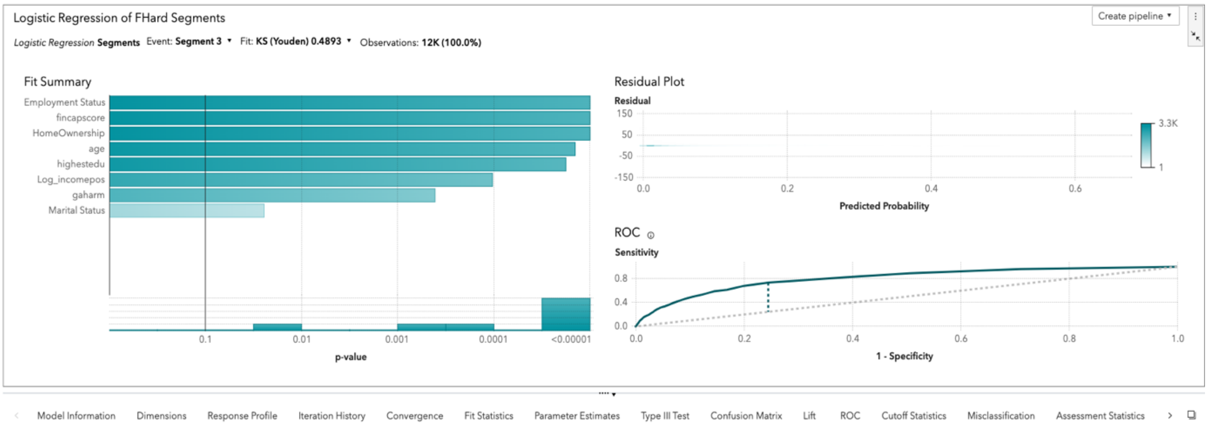
Parameter	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	-1.17111	0.241415	23.53257	<0.00001
Employment Status Employed	-0.46397	0.059568	60.6659	<0.00001
Employment Status Not Employed	0	.	.	.
HomeOwnership No Home Own	0.364919	0.055097	43.86689	<0.00001
HomeOwnership Own Home	0	.	.	.
Marital Status Married	-0.3532	0.062032	32.41991	<0.00001
Marital Status Not Married	0	.	.	.
age	0.007552	0.001993	14.36216	0.00015
gaharm	0.040622	0.017878	5.162944	0.02307
fincapscore	-0.17269	0.020631	70.06665	<0.00001
highestedu	-0.04769	0.016478	8.376723	0.00380
Log_incomepos	-0.00543	0.025823	0.044285	0.83332

Key Variables, Direction, Strength and Significance:

- Employment Status (Employed): Negative coefficient (-0.464, highly significant)
- Homeownership (No Homeownership): Positive coefficient (0.365, highly significant)
- Marital Status (Married): Negative coefficient (-0.353, highly significant)
- Age: Positive coefficient (0.0076, highly significant)
- Gambling Harm (gaharm): Positive coefficient (0.041, significant)
- Financial Capability (fincapscore): Negative coefficient (-0.173, highly significant)
- Education (highestedu): Negative coefficient (-0.0477, highly significant)

Employment and homeownership act as major protective factors against extreme financial distress in this segment, with non-homeowners and unemployed individuals more likely to be in financial trouble. Older age, gambling harm, lower financial capability, and lower education significantly increase the vulnerability and likelihood of being in this segment, reflecting the compounding impact of structural disadvantages and financial mismanagement. Addressing financial literacy and employment opportunities may help alleviate this group's vulnerability.

Image 6: Segment 3 – Seeking Help to Escape Distress – Logistic Regression (Probit)



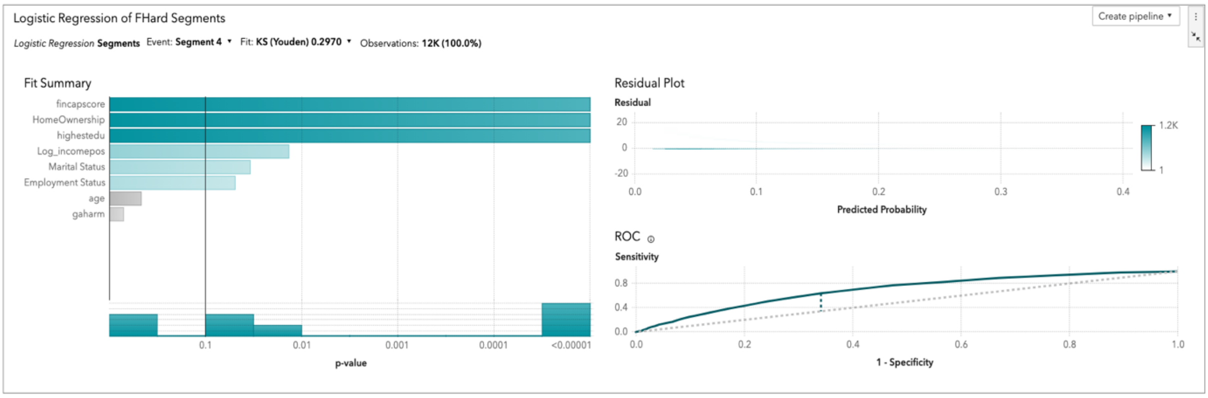
Parameter	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	-1.26147	0.225417	31.31686	<0.00001
Employment Status Employed	-0.63646	0.052645	146.1642	<0.00001
Employment Status Not Employed	0	.	.	.
HomeOwnership No Home Own	0.45323	0.048681	86.68027	<0.00001
HomeOwnership Own Home	0	.	.	.
Marital Status Married	-0.12085	0.053738	5.05755	0.02452
Marital Status Not Married	0	.	.	.
age	-0.00819	0.001888	18.83284	0.00001
gaharm	0.056128	0.015882	12.48907	0.00041
fincapscore	-0.20933	0.018431	128.9898	<0.00001
highestedu	-0.0635	0.014802	18.40597	0.00002
Log_incomepos	0.094381	0.024306	15.07835	0.00010

Key Variables, Direction, Strength and Significance:

- Employment Status (Employed): Negative coefficient (-0.636, highly significant)
- Homeownership (No Homeownership): Positive coefficient (0.453, highly significant)
- Marital Status (Married): Negative coefficient (-0.121, significant)
- Age: Negative coefficient (-0.0082, highly significant)
- Gambling Harm (gaharm): Positive coefficient (0.0561, highly significant)
- Financial Capability (fincapscore): Negative coefficient (-0.209, highly significant)
- Log of Income: Positive coefficient (0.0944, highly significant)

This segment consists of individuals trying to regain financial stability. Individuals who do not own their home, younger individuals, and those with lower financial capability and education are more likely to belong to this segment. Although they may be employed, their overall financial literacy and the presence of gambling harm continue to challenge their efforts to escape distress. Income has a slight protective effect, but it's not enough to overcome the other factors. Helping these individuals improve financial capability and reducing gambling harm could assist in their recovery.

Image 7: Segment 4 – Responding to Financial Shock – Logistic Regression (Probit)



Model Information | Dimensions | Response Profile | Iteration History | Convergence | Fit Statistics | Parameter Estimates | Type III Test | Confusion Matrix | Lift | ROC | Cutoff Statistics | Misclassification | Assessment Statistics

Parameter	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	-1.18748	0.175156	45.96243	<0.00001
Employment Status Employed	0.092152	0.04684	3.870532	0.04914
Employment Status Not Employed	0	.	.	.
HomeOwnership No Home Own	0.384234	0.037636	104.2254	<0.00001
HomeOwnership Own Home	0	.	.	.
Marital Status Married	0.082677	0.039047	4.483248	0.03423
Marital Status Not Married	0	.	.	.
age	-0.00106	0.001447	0.534272	0.46482
gaharm	-0.00612	0.016404	0.139103	0.70917
fincapscore	-0.20721	0.014441	205.8742	<0.00001
highestedu	-0.05281	0.010695	24.38763	<0.00001
Log_incomepos	0.045756	0.018537	6.093005	0.01357

Key Variables, Direction, Strength and Significance:

- Employment Status (Employed): Positive coefficient (0.092, significant)
- Homeownership (No Homeownership): Positive coefficient (0.384, highly significant)
- Marital Status (Married): Positive coefficient (0.083, significant)
- Financial Capability (fincapscore): Negative coefficient (-0.207, highly significant)
- Log of Income: Positive coefficient (0.0458, significant)
- Education (highestedu): Negative coefficient (-0.0528, highly significant)

Individuals in this segment are likely responding to unexpected financial shocks. While employment and income act as protective factors, low financial capability and education continue to put them in a vulnerable position.

Non-homeowners are also more likely to be in this group, highlighting the impact of housing insecurity in the scenario of financial shocks. Married individuals are more likely to belong to this group, potentially reflecting family related financial responsibilities. Addressing financial literacy and education gaps could help these individuals navigate and recover from sudden financial shocks to their finances.

The segmentation and regression analysis reveals several similarities and distinctive differences across the four segments. Employment and home ownership consistently emerge as crucial protective factors against financial distress, as individuals who are employed and own homes are less likely to experience financial hardship. Financial capability also plays a pivotal role in determining the likelihood of financial distress, with lower financial literacy contributing to greater financial vulnerability in all segments except for Segment 1, where financial capability has a positive impact. Similarly, marital status generally offers more financial stability, with married individuals being less likely to face distress in most segments, although Segment 3 presents an exception where being married has a negative effect, potentially due to those with gambling behaviour and criticism from the partner.

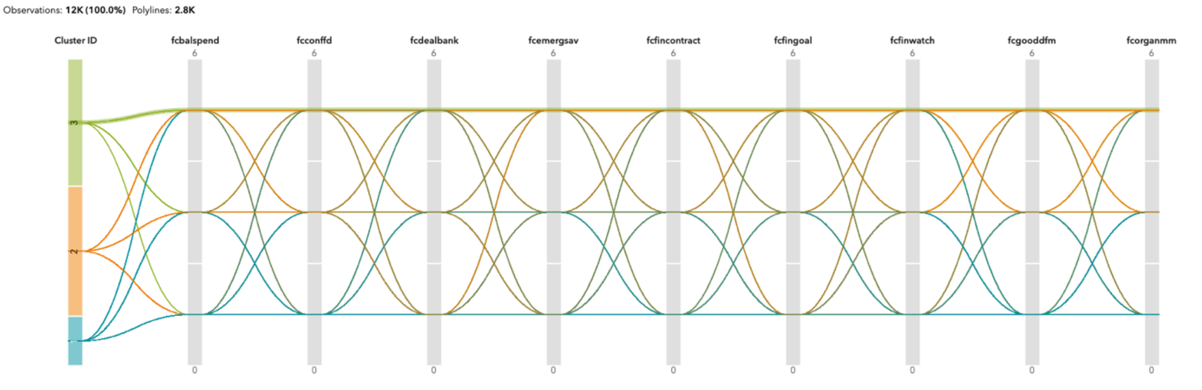
There are key differences between the segments. Segment 1 stands out as the only group where financial capability has a positive influence, indicating that individuals in this segment possess strong financial management skills and enjoy stability. Segment 2 faces the most extreme vulnerability, where older age, unemployment, and non-homeownership are the dominant factors contributing to financial distress. In contrast, Segment 3 is characterised by younger individuals who, despite being employed, continue to face financial challenges due to low financial capability and issues like gambling harm. Lastly, Segment 4 represents individuals responding to unexpected financial shocks. While employment and income offer some protection, these individuals are still vulnerable due to poor financial management and housing insecurity.

In conclusion, while some factors such as employment and homeownership play a protective role across all segments, financial capability provided an insightful discrimination between financially stable and distressed segments. Also, while income does influence financial hardship, analysis show that high income individuals may still struggle due to factors like high debt, unexpected financial shocks, or gambling behaviour. The specific drivers of financial distress differ in each case, requiring targeted interventions to address the unique challenges faced by each group, and financial literacy could be the key to minimise financial hardship.

3.2 Clustering Financial Capability

In order to get a better understanding of the dynamics and behaviours of individuals I decided to cluster financial capability and gambling harm. Both SAS Miner and SAS Viya clearly clustered and segmented them into low, medium and high (Image 9). Note that overall, financial capability is generally medium-high (clusters 2 and 3 have over 5000 observations, while cluster 1 has 2000), so we could potentially have just 2 clusters, low and high financial capability.

Image 9: Financial Capability Clusters (SAS Viya)



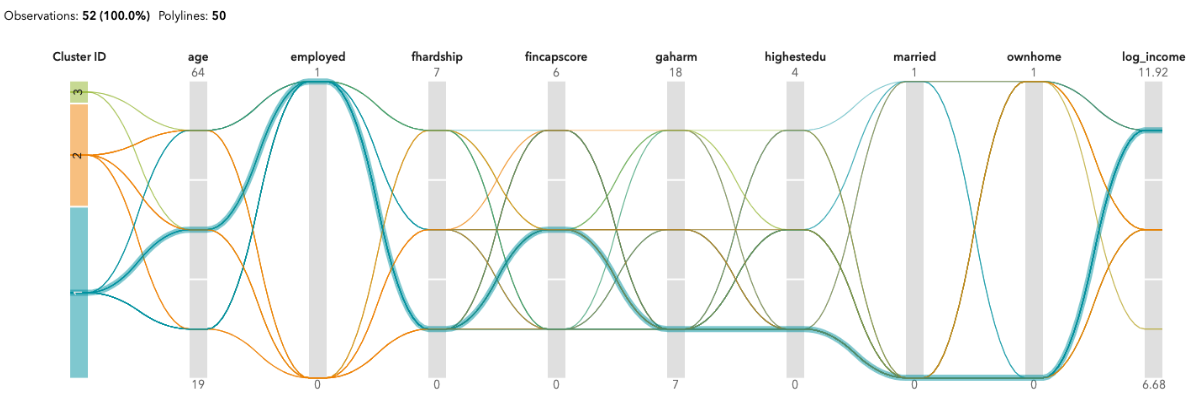
Centroids | Cluster Summary | Model Information | Within Cluster Statistics | Iteration History | Standardization | Interval information | Parallel Coordinates Plot

Cluster ID	fcbalspend	fcconffd	fcdealbank	fcemergsav	fcfincontract	fcfingoal	fcfinwatch	fcgooddfm	fcorganmm
1	1.524	2.32	1.7945	1.6545	2.9575	1.6465	2.061	1.8285	1.5555
2	3.297503	3.670669	3.359634	3.399085	4.316752	3.129217	3.785401	3.683629	3.220888
3	4.989492	4.948044	4.951936	5.087566	5.414478	4.685931	5.145554	5.241292	4.944736

3.3 Clustering Medium-High Gambling Harm Individuals

In the case of gambling behaviour, even though they fortunately represent a small percentage of individuals, I dived a bit deeper focusing only on individuals that displayed medium to high gambling harm behaviour (gaharm score of 7 or higher). Thus, I filter these individuals to cluster them into 3 groups to understand better who are they likely to be and their dynamics, while being aware that taking into account that the small number of observations (52 individuals) means these results may not generalise (Image 10).

Image 10: Clustering Medium-High Gambling Harm Individuals



Centroids | Cluster Summary | Model Information | Within Cluster Statistics | Iteration History | Standardization | Interval information | Parallel Coordinates Plot

Cluster ID	age	employed	fhardship	fincapscore	gaharm	highestedu	married	ownhome	log_income
1	36.16667	1	1.466667	2.351852	9.283333	1.2	0.166667	0.133333	11.01
2	43.38889	0.166667	2.444444	3.209877	9.972222	0.666667	0.055556	0.5	9.68
3	50.25	0.5	5.75	1.611111	14.75	2.5	9.84E-14	0.75	9.10

Bearing in mind the limitations, the clustering results indicate that medium-high gamblers could potentially be split into 3:

- Cluster 1 (30 observations): "*Young Gamblers*", likely represents younger, single who do not own property, but are employed individuals with relatively high financial capability. Despite experiencing some financial hardship, they manage their situation fairly well, likely due to employment, above median income and better financial skills.
- Cluster 2 (18 obs.): "*Middle-Age Crisis Gamblers*", potentially consists of struggling middle-aged individuals, likely unemployed individuals with high financial capability but facing moderate financial hardship. They have the lowest education level, are mostly single, have mixed homeownership, while having less available income.
- Cluster 3 (4 obs.): "*Addicted Gamblers*", seems to represent reckless gambling addicts, older individuals with a mix of employment statuses. Despite having higher education and more homeownership, they tend to have low financial capability, income and suffer from the highest level of gambling harm and financial hardship. Their situation may reflect the impact of gambling on their financial stability.

4. Conclusions

The objective of this analysis was to cluster the population into segments based on financial attitudes and behaviours to better inform authorities of strategies for assisting those facing, or at risk of falling into, financial distress. The analysis revealed that financial hardship is not solely a problem for low income groups, as high income individuals may also in financial hardship due to high debt, unexpected financial shocks, and gambling behaviour. These findings highlight the need for tailored financial literacy advice and early intervention, particularly for low-income households vulnerable to debt. The aim is to promote more targeted financial literacy programmes that enhance people's financial and debt management skills while also mitigating the risks posed by harmful gambling habits.

Additionally, the analysis showed that employment and homeownership consistently act as protective factors, reducing the likelihood of financial hardship across all segments. On the other hand, low financial capability emerged as a key driver of vulnerability, regardless of income level. This reveals the importance of improving financial literacy to build resilience, especially for those struggling with gambling issues or unexpected financial shocks.

By understanding the distinct characteristics of each segment, more effective and targeted interventions can be developed. For example, programmes focusing on budgeting, saving, and responsible credit use, whereas those facing unemployment could be supported through debt counselling and financial aid resources.

In conclusion, by implementing these strategies, authorities and financial institutions can help improve financial resilience among individuals, reducing the risks of financial hardship and promoting better financial wellbeing overall.

5. References

Armstrong, A. R., & Thomas, A. (2017). Gambling participation, expenditure, and risk of harm in Australia, 1997–1998 and 2010–2011. *Journal of Gambling Studies, 34*(1), 255–274.

Bourova, E., *et al.* (2018). Impacts of financial literacy and confidence on the severity of financial hardship in Australia. *Australasian Accounting, Business and Finance Journal, 12*(4), 3-25.

Collica, R. S. (2021). *Segmentation analytics with SAS Viya: An approach to clustering and visualization*. SAS Institute.

Koomson, I., Awaworyi Churchill, S., & Munyanyi, M. E. (2022). Gambling and financial stress. *Social Indicators Research, 163*(2), 473-503.

Marko, S., *et al.* (2023). The lived experience of financial harm from gambling in Australia. *Health Promotion International, 38*(1), 1-13.

Scroggins et al., (2006). *Borrowing Segmentation in and Credit Scoring in Bank Consumer Lending*. Business Quest. State University of West Georgia

Worthington, A. C. (2006). Predicting financial literacy in Australia. *Financial Services Review, 15*(1), 59-79.