Table of Contents

Executive summary

This report analyses and forecasts key variables that drive or hinder total voluntary super contribution (TVSC) and application satisfaction in order to provide informed recommendations. Descriptive analytic techniques were used to identify key variables that would impact TVSC and Application satisfaction (AS). Predictive analytic techniques were then used to forecast how certain variables impact TVSC and application satisfaction.

Based on insights derived from analysis, it is recommended that Moneysoft should seek to target underperforming groups within the select variables to ensure maximum improvement in TVSC and AS. The team encourages MoneySoft to provide incentive schemes, user testimonials, personalised tutorials and access to financial advice.

Introduction

Report aims to explore issues impacting user superannuation contributions and overall application satisfaction on the Roundups run by Moneysoft. This issue is prevalent within the wider Australian landscape as "superannuation amount is insufficient to sustain an adequate retirement" *(Deloitte, 2014).*

The key factors associated with TVSC includes the underperformance in contribution from low income individuals as well as individuals with low financial literacy who have limited understanding of superannuation investments. The factors associated with application satisfaction includes the large % of users not watching the app tutorial, hindering their understanding of app functionality. Additionally the application satisfaction is further compromised from the lack of users receiving adequate personal financial advice from super firms. This is a key business issue that impacts how the user manages the app and overall value they gain from RoundUps. There is a low level of contribution within the Australian context due to lack of interest and capability *(Bateman, 2014)*, however this report aims to advise moneysoft how they can remedy these issues.

Data analysis - Descriptive

Total Voluntary Super Contribution

A scatterplot was used to show the strength of the correlation between numerical variables, TVSC and monthly income, as seen in *Figure 1*. TVSC variable shows the total amount of funds contributed towards the roundups application by each user, and monthly income represents an individual's earnings for each month.

An analysis by the Economic Society of Australia revealed that TVSC are "strongly increasing in wages" and that the "elasticity of voluntary contributions to wages is >1" *(Bateman, 2014)*. These variables can additionally be used in predictive analysis to determine the impact monthly income has on TVSC. Figure 1 depicts a strong positive correlation of 0.792 as TVSC is incrementally increasing with individual monthly income. Moreover, the positive gradient of the LOBF provides insight into the positive correlation between the two variables.

Figure 1: *Scatter plot of monthly income VS voluntary super contribution*

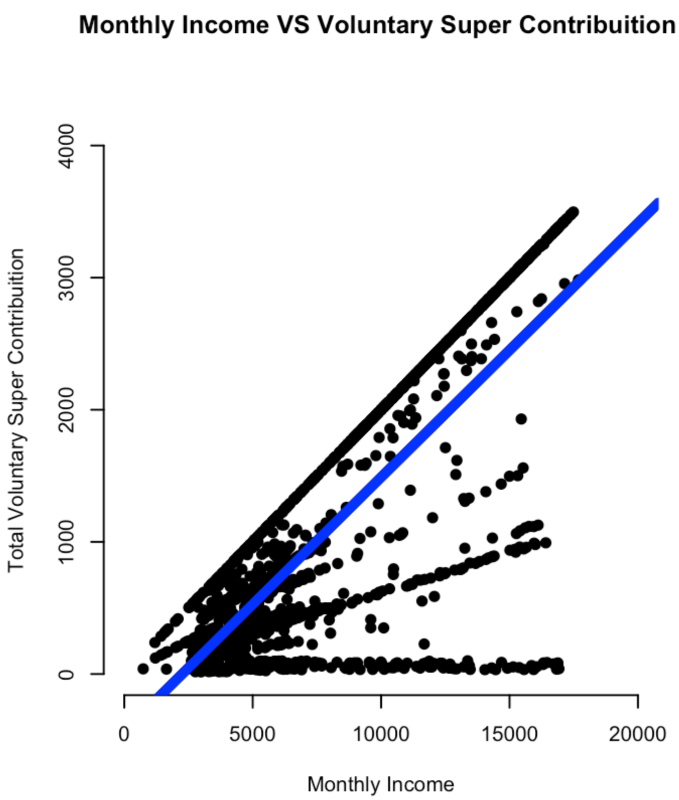
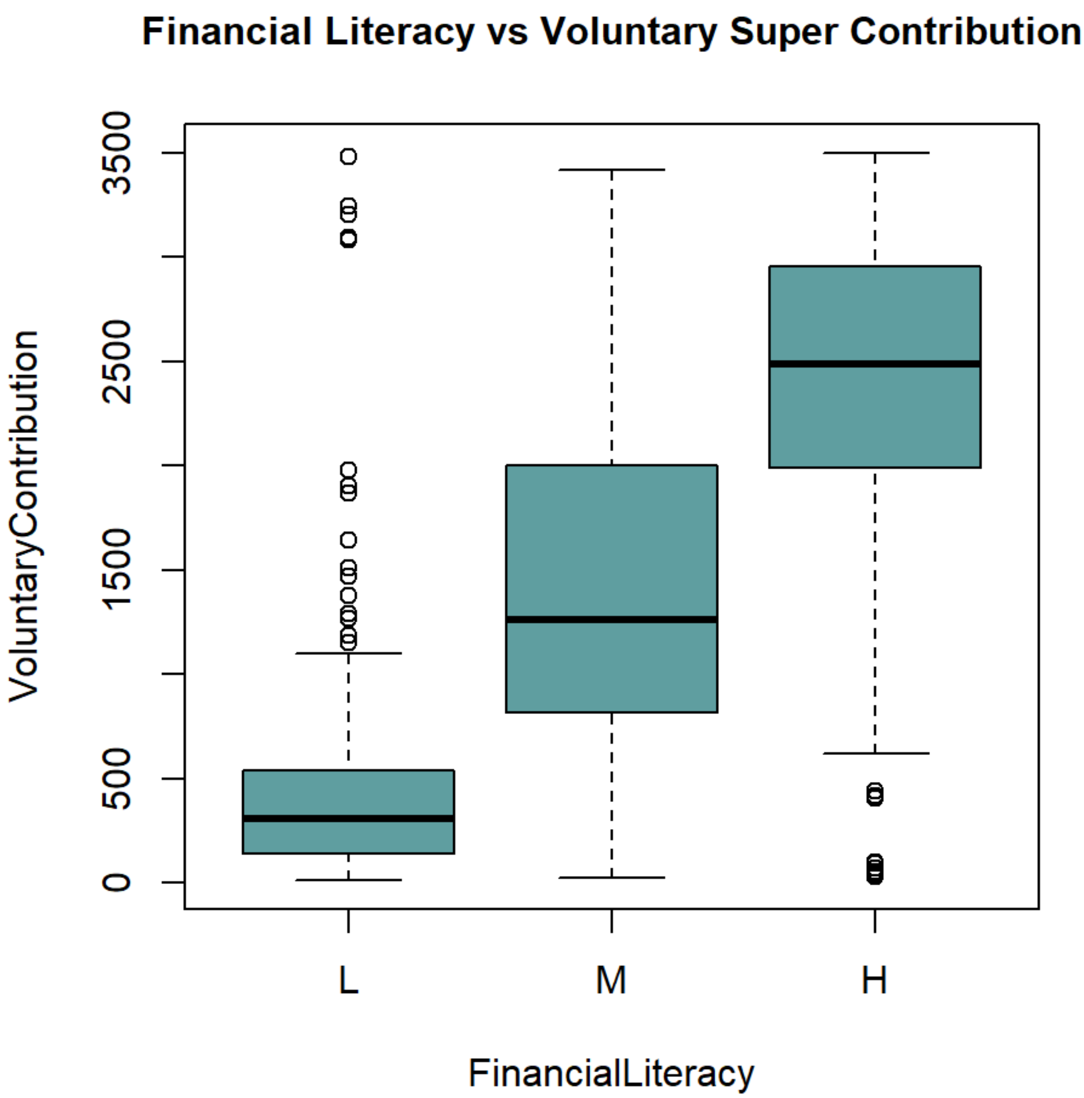


Figure 2 effectively utilises a boxplot to showcase the distribution of the TVSC against the categorical financial literacy variable. The negatively skewed data illustrates that a significant distribution of the values is likely to invest in their superannuation voluntarily. The median propensity to TVSC when financial literacy is low is less than when an individual has a high financial literacy. Additionally, removing outliers for the values with low financial literacy further indicates that a higher financial literacy leads to a higher propensity to contribute to superannuation. The variables of TVSC and financial literacy were utilised in this technique as higher levels of

financial literacy, including self-rated and simple literacy, have been shown to aid individuals in investment choices, including superannuation *(Palm, 2014).* Hence, financial literacy is a suitable variable for predictive analysis.

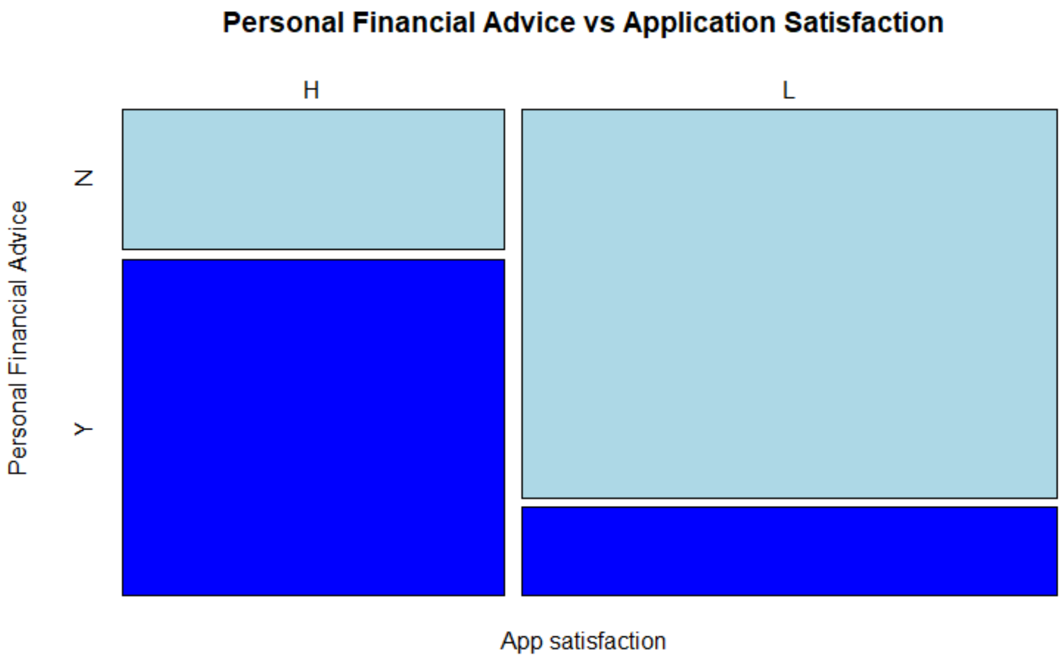
Figure 2: *Boxplot of Financial VS voluntary super contribution*



Application satisfaction

While several factors can drive application satisfaction, Wyzowl study indicates that 72% of people prefer to use videos to learn about a product or service *(Povidom, 2022).* Thus, MoneySoft's application tutorial should directly impact customers' application satisfaction. *Figure 3* depicts a stacked barplot to identify whether watching the application tutorial impacted application satisfaction. The stacked barplot was chosen as it clearly illustrates the proportional difference in high or low application satisfaction of the categorical variable app show tutorial. Specifically indicated in *Figure 3*, 25% of the population had high satisfaction without watching the tutorial, while 85% of those who watched the tutorial reported high satisfaction. This is because users have a better experience with the application when they understand its functionalities. A strong correlation exists between watching the tutorial and high application satisfaction and vice versa. Hence, the app shows that tutorial is identified as a suitable variable for predicting future outcomes using predictive analysis.

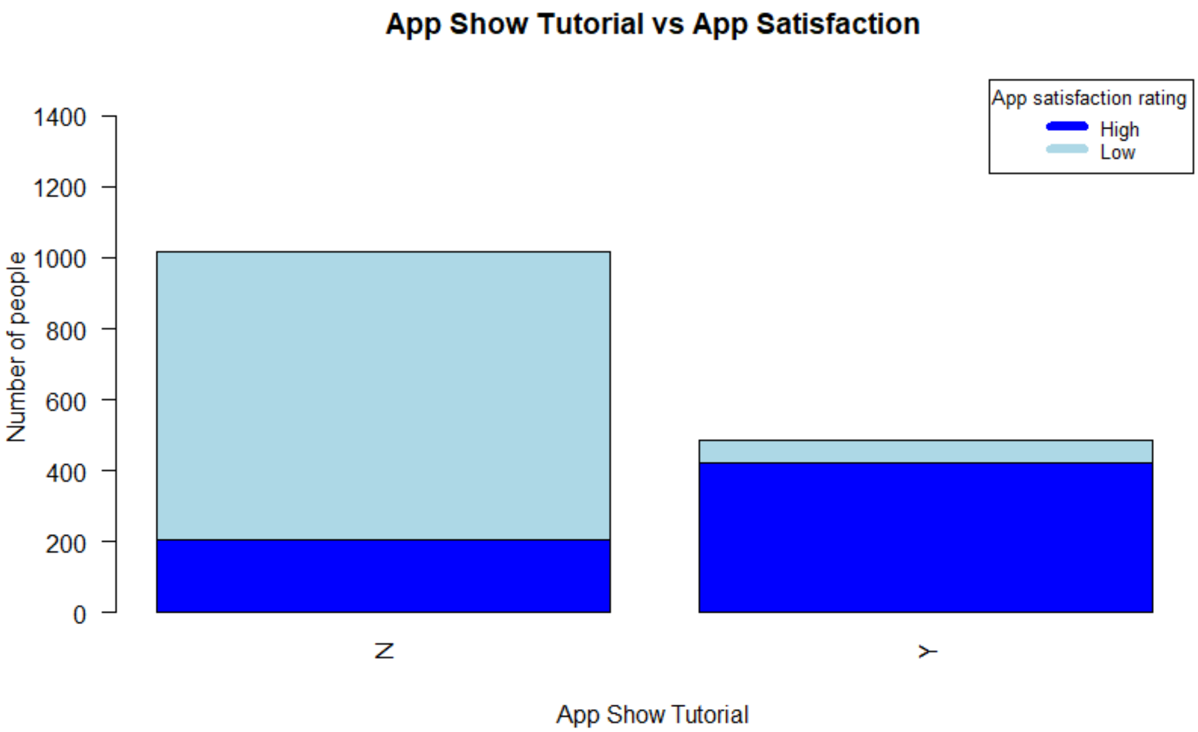
Figure 3: *Stacked barplot of Application satisfaction vs App Show Tutorial*



Furthermore, as discussed above, financial literacy positively impacts TVSC, inferring that personal financial advice may impact application satisfaction as financial literacy can be improved by seeking advice (Janine Mace, 2021). Thus, a mosaic plot has been used to assess whether the categorical variable personal financial advice impacts user application satisfaction. In addition, a mosaic plot provides insight into the proportionality of the two categories, high and low - comparing them directly.

As indicated in *Figure 4,* a significantly larger proportion of users who watched the tutorial reported having high satisfaction compared to the population that had not watched the tutorial. As personal financial advice enhances an individual's understanding of their financial situation, they are more likely to derive value from using MoneySoft's roundup application, thus experiencing higher satisfaction. Hence, personal financial advice is also suitable for predictive analysis as it demonstrates the impact on AS through historical data.

Figure 4: *Mosaic plot of application satisfaction vs Personal Financial Advice*



Predictive

Predictive Technique 1
In addition to the findings of descriptive analytics, multiple linear regression modelling techniques were utilised to determine the predictive capacity of test variables in finding the impact of TVSC. This modelling technique was utilised due to its efficacy in predicting single dependent variables such as TVSC, via independent variables whose values are known - as seen in table 5.

Following descriptive analytics and industry trends, the variables C_MonthlyIncome and C_ FinancialLiteracy were found to have the most significant impact on TVSC.
Many variables seen in table 2 were deemed
d insignificant to impact TVSC since they exceeded the p-value of 0.05. However, the select variables MonthlyIncome and FinancialLiteracy can be interpreted to be the most statistically significant from zero as they hold a p-value of <0.05, ultimately rejecting the null hypothesis. Their considerable significance can also be noted through the labelling using "***".

Multiple R squared and adjusted R squared can be used to determine how well the model fits the data. For example, the simple linear regression, as seen in Tables 3.1 and 3.2 for variables C_MonthlyIncome and C_ FinancialLiteracy, depicts respective multiple R squared values of 0.607 and 0.643. This is represented in Figures 5.1 and 5.2, demonstrating a strong positive correlation between the independent and dependent variables.

Figure 5.1: *Scatter plot of Monthly Income VS Voluntary Super Contribution*

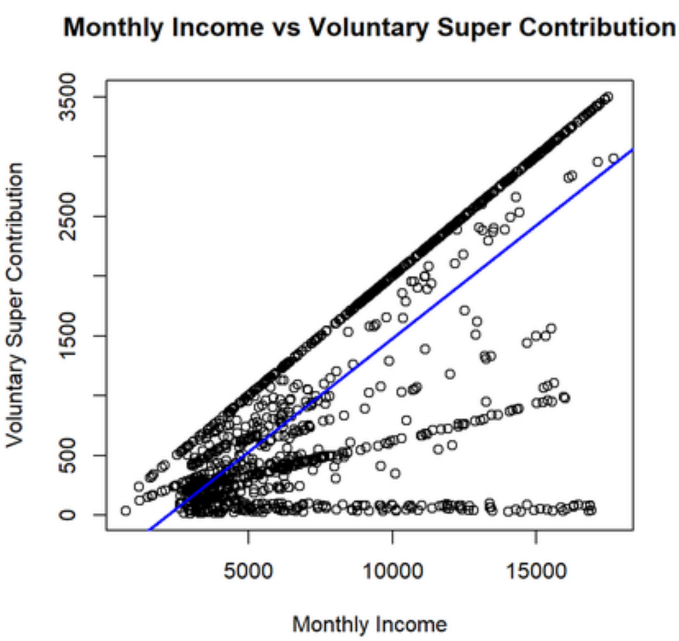
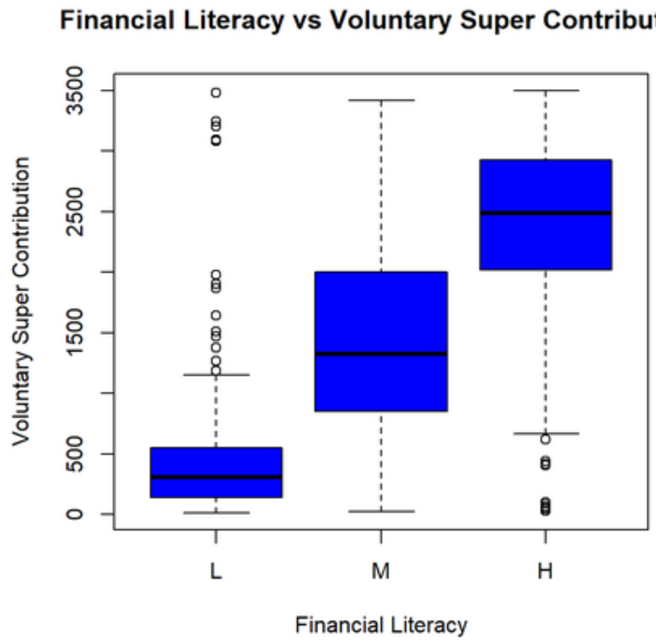
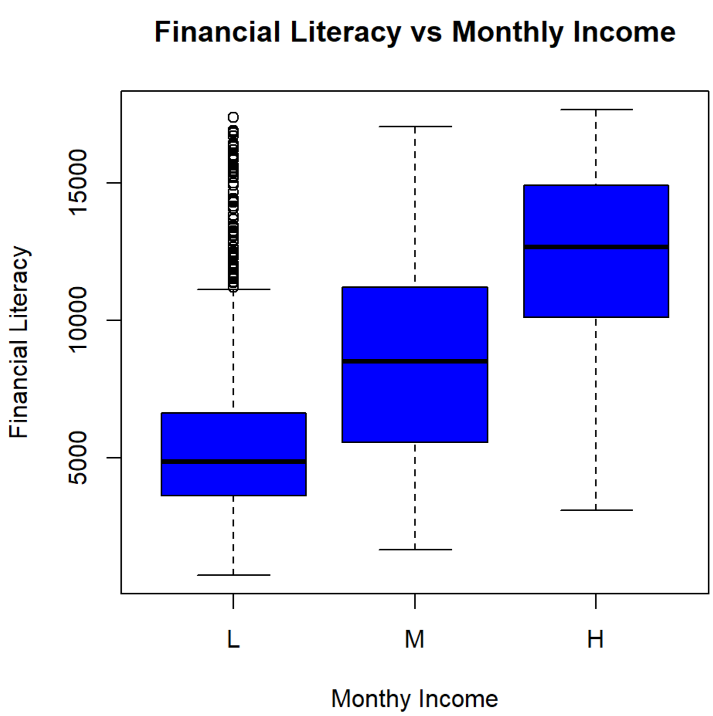


Figure 5.2: *Box plot of Monthly Income VS Voluntary Super Contribution*



When this is compared against the multiple linear regression model of the variables C_MonthlyIncome and C_ FinancialLiteracy in table 3.3, the adjusted R squared is 0.766. The model is further validated in Figure 5.3 whereby a strong positive trend between variables is seen. Thus, 76% of variability in TVSC can be explained by MLR. This high adjusted R squared value symbolises the best fit between the model and data.

Figure 5.3: *Box plot of Voluntary Superannuation Contribution as a function of Monthly Income and Financial Literacy*



Predictive Technique 2

To forecast how certain factors were impacting the app satisfaction of Roundups' users, logistic regression was applied with multiple variables. Logistic regression is a predictive technique that enables the prediction of binary outcomes making it ideal in examining how certain variables impact the binary outcome of high and low app satisfaction. To explore the impact of various independent variables influencing app satisfaction, three logistic regression models were created with pairs of differing predictors. Each models' performance was then assessed using an accuracy table and ultimately the strongest model was chosen to present. Models 2 and 3 (Figures 7 and 8 in the Appendix) were both discarded as the accuracy of both models was below 70%. Furthermore, the misclassification rate for both models was greater than 30%, implying a weak accuracy in the results and the weak correlation of the predictors used.

However, model 1 demonstrated a strong correlation between the predictors and app satisfaction. The two predictors used in model 1 were App_ShowTutorial and C_PersonalFinancialAdvice. The summary table in the appendix (Table 1 in Appendix) showcases the results of applying logistic regression to all the variables. it is evident that for both App_ShowTutorial and C_PersonalFinancialAdvice the p-values are significantly lower than 0.05. This signifies that the null hypothesis can be rejected and that both predictors have a significant correlation to app satisfaction. The coefficients of both predictors were greater than 0, indicating a positive correlation with app satisfaction. This suggests that users who receive personal financial advice and watch the app tutorial, are more likely to have higher app satisfaction.

Model Performance

The performance of both predictive models were analysed to assess the accuracy of the correlations found in data analysis. 80% of the data was used to train the model and 20% of the data was used to test the model.

The mean-square error of the variables monthly income and financial literacy seen within the multiple linear regression model was computed using the validations set approach. Table 4 demonstrates a MSE of 247336.2 for monthly income and financial literacy compared to an error of 490616.3 for age, gender and relationship status variables. The low-MSE of monthly income and financial literacy variables forecasts strong predictive performance in determining TVSC impact compared to age, gender and relationship status variables.

In light of the logistic regression model, figures 6.1 and 6.2 represent the confusion matrix formed for both data sets. Moreover, figure 7 represents the accuracy table for the trained and test confusion matrix. The 82.5% accuracy rate validates the strong relationship between the two variables and application satisfaction. Furthermore, the accuracy of low app satisfaction for both matrices was greater than 90%, providing significant validation for the correlation between the predictors and low AS. Although the accuracy of the model with high app satisfaction is low, this can be explained by other predictors that may have a stronger correlation to high app satisfaction than the predictors used. Therefore, the correlation determined by the model was accurate.

Figure 6.1 *Trained Confusion Matrix for Logistic Regression Model 1 (Personal Financial Advice & App Show Tutorial)*

Predicted Values \ Actual Values	H	L
H	332	50
L	166	652

Figure 6.2: Test Confusion Matrix *for Logistic Regression Model 1 (Personal Financial Advice & App Show Tutorial)*

Predicted Values \ Actual Values	H	L
H	89	13
L	38	160

Figure 6.3: *Accuracy Table for Logistic Regression Model 1*

	Trained Data	Test Data
Accuracy of model	(332+652)/1200 = 0.820 or 82.0% Accurate	(89+160)/(300) = 0.83 or 83% Accurate
Accuracy of High App Satisfaction	332/(332+166) = 0.666 or 66.6% Accurate	89/(89+38) = 0.701 or 70.1% Accurate
Accuracy of Low App Satisfaction	652/(652+50) = 0.929 or 92.9% Accurate	160(160+13) or 0.925 or 92.5% Accurate
Precision	332/(332+50) = 0.869 or 86.9% Precision	89/(89+13) = 0.873 or 87.3% Precision
Misclassification	(50+70)/1200 = 0.180 or 18.0% Error	(13+38)/300 = 0.17 or 17% Error

Recommendations

Moneysoft can improve the business problem of the TVSC by addressing low-income users. The data analysis depicts a direct trend between an individual's income and contribution; hence providing low-income users access to incentive schemes will boost their morale to make TVSC Moneysoft can leverage customers to access government incentive schemes such as the Australian Tax Office offering a super co-contribution, which enables individuals to boost their total savings *(ATO, 2019)*.

Furthermore, Moneysoft can increase its TVSC by destigmatising TVSC by showing user testimonies. Individuals with low financial literacy are daunted by the process of voluntary contributions as the results of TVSC are unknown and therefore seem risky. Thus, Moneysoft can provide financially unaware individuals an easy route to TVSC*)*. Moneysoft can implement a scheme in which successful application users can provide testimonies as a form of social proof that the contribution is beneficial to their future. Many users do not feel comfortable with just company marketing as it is biassed *(Sonnenberg, 2022).*

Application satisfaction is derived from factors such as ease of use and functionality as these allow the user to achieve their purpose (*Shah, 2021*). MoneySoft could make the tutorial compulsory for new users, ensuring they understand the app's functionality. To increase engagement of feature usage, the tutorials should be personally curated for each user to increase the probability of feature usage. For example, tutorials that recommend specific settings based on monthly income should be created so that more users from different income brackets will have a higher AS.

Finally, the analysis indicates that receiving financial advice is essential for AS. Thus, MoneySoft should aim to provide customers with the access to financial resources that can improve their financial literacy. This will allow users to better understand their financial situation, meaning they are more likely to find value through using the application and thus experience a higher AS.

Ethics

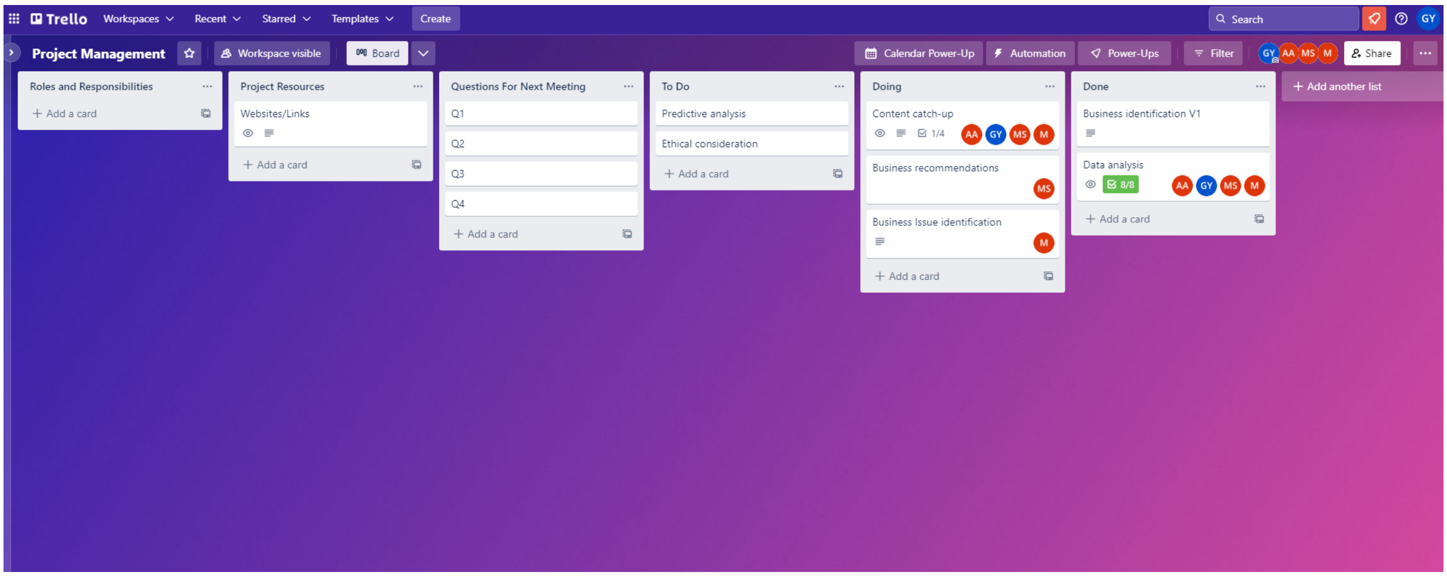
Ethical Consideration	**Recommendations**
User's privacy	Data-driven business decisions are essential for current business operations *(Lee, 2016)*. However, data utilised to acquire business insights contains sensitive demographic and identity data concerning the application's users, which users must consent to. MoneySoft must provide users with the means to report and provide feedback on how protected they feel. MoneySoft must act in adherence to Australian Privacy Principles. Users prefer unlinkability, where they can use MoneySoft's services without being tracked. ThusMoneySoft can have its data proctored using pseudo identities.
Data Security	Data security entails privacy breaches that can dilute users' trust in MoneySoft's application. In addition, consumers are concerned with the misuse of data in the case of a breach in security where their data can be sold to third parties, putting users at risk of cybercrime. As a result, MoneySoft must embed mechanisms such as authorisation of the user, which can restrict actions taken on the data.
Confirmation bias	Confirmation bias is when analysts hold previous beliefs to explain patterns in the dataset and nitpick specific data to confirm them. This is apparent in how social media manipulates its algorithm to cater content towards users' prior beliefs *(Team, 2021)*. Therefore, MoneySoft's analysts must consider alternate hypothesis once the data has been analysed.

Project management

Planning

To display a clear timeline, a Gantt Chart was created and shared amongst members via google sheets where deadlines, milestones and responsibilities were stated.

In order to boost team morale and team chemistry, milestone strategies such as weekly zoom work sessions and team socials were implemented. Furthermore, a trello board was created to record down actionables and ensure all tasks were complete.



Finally, meeting minutes are recorded through microsoft word where each member takes turns taking meeting minutes, allowing members to have a clear direction of the project.

Reflection

The use of project management tools such as Gantt chart and Trello board assisted the team in completing a high-quality report.

One attribute of the team culture that assisted with both the quality of the report and the collaboration process was having an open-minded mentality. Specifically, each member's input was discussed and valued, which meant ideas were synthesised into a better product, increasing overall coherency.

However, group members should conduct more regular updates so team members are more aware of the progress. As the team faced an issue of updating team members after their absence. Thus, for future improvements, a team member should have a separate meeting with absent members and update them on the progress to ensure that correct information is communicated.

References

Admin (2022) *Why tutorial videos are so important for SAAS businesses!*, *Video Production Services - Contact*. Available at: https://www.povidom.com/news-2/why-tutorial-videos-are-so-important-for-saas-businesses/ (Accessed: November 4, 2022).

ATO, 2019. ‘Government Super Contributions’. Australian Taxation Office. Last Accessed 6th October. < https://www.ato.gov.au/individuals/super/growing-your- super/adding-to-your-super/government-super-contributions/#:~:text=The%20super%20co%2Dcontribution%20helps,a%20maximu m%20amount%20of%20%24500> (Accessed: November 2, 2022).

Bateman H, Deetlefs J, Cepar L, Newell B, Ortmann A, Thorp S, June 2014. ‘Just Interested or Getting Involved? An Analysis of Superannuation Attitudes and Actions’. Available at: The Economic Society of Australia.<https://onlinelibrary-wiley-com.wwwproxy1.library.unsw.edu.au/doi/pdfdirect/10.1111/1475-4932.12107> (Accessed: November 1, 2022).

(2014) *Deloitte*. Available at: https://www2.deloitte.com/content/dam/Deloitte/au/Documents/financial-services/deloitte-au-fs-adequacy-australian-superannuation-system-0614.pdf (Accessed: 2022).

Lee, W. (2016) *An ethical approach to data privacy protection*, *ISACA*. ISACA. Available at: https://www.isaca.org/resources/isaca-journal/issues/2016/volume-6/an-ethical-approach-to-data-privacy-protection#:~:text=Privacy%20breaches%20disturb%20trust%20and,legal%20right%20to%20the%20data. (Accessed: November 3, 2022).

Mace, J. (2022) *Why financial literacy is vital for a good retirement*, *SuperGuide*. Available at: https://www.superguide.com.au/retirement-planning/why-financial-literacy-is-vital-for-a-good-retirement (Accessed: November 4, 2022).

Palm, C. (2014) *Financial literacy and superannuation investment decision-making in a choice environment: An exploratory study*. Queensland University of Technology. Available at: https://eprints.qut.edu.au/78464/4/78464.pdf (Accessed: October 5, 2022).

Sonnenberg, A. (2022) *The importance of using testimonials*, *Content Marketing Agency | Content Marketing Services by CopyPress*. Available at: https://www.copypress.com/blog/importance-using-testimonials/#:~:text=Testimonials%20help%20build%20your%20brand's,leadership%20toots%20its%20own%20horn. (Accessed: November 4, 2022).

Team, T.M. (2021) *The 6 most common types of bias when working with data*, *Metabase*. Available at: https://www.metabase.com/blog/6-most-common-type-of-data-bias-in-data-analysis (Accessed: November 3, 2022).

What makes a good app? 10 key elements to develop a successful app (2021) *Techuz Blog*. Available at: https://www.techuz.com/blog/what-makes-a-good-app/#:~:text=A%20good%20app%20is%20a,user%20retention%20and%20generates%20revenue. (Accessed: November 4, 2022).

Whitney, L. *et al.* (2022) *Data privacy is a growing concern for more consumers*, *TechRepublic*. Available at: https://www.techrepublic.com/article/data-privacy-is-a-growing-concern-for-more-consumers/ (Accessed: November 3, 2022).

Appendix

Figure 7.1: *Trained Confusion Matrix for Logistic Regression Model 2 (Relationship Status & Gender)*

Actual Values

Predicted Values

	H	L
H	164	117
L	334	585

Figure 7.2: *Test Confusion Matrix for Logistic Regression Model 2 (Relationship Status & Gender)*

Actual Values

Predicted Values

	H	L
H	39	35
L	88	138

Figure 7.3: *Accuracy Table for Logistic Regression Model 2 (Relationship Status & Gender)*

	Trained Data	Test Data
Accuracy of model	(164+585)/1200 = 0.624 or 62.4% Accurate	(39+138)/300 = 0.590 or 59.0% Accurate
Accuracy of High App Satisfaction	164/(164+334). = 0.329 or 32.9% Accurate	39/(39+88) = 0.307 or 30.7% Accurate
Accuracy of Low App Satisfaction	585/(585+117) = 0.833 or 83.3% Accurate	138/(138 + 35) = 0.798 or 79.8% Accurate
Precision	164/(164+117). = 0.584 or 58.4% Precision	39/(39+35) = 0.527 or 52.7% Precision
Misclassification	(334+117)/1200 = 0.376 or 37.6% Error	(88+35)/300 = 0.410 or 41.0% Error

Figure 8.1: *Trained Confusion Matrix for Logistic Regression Model 3 (Home Ownership & Employment)*

Actual Values

Predicted Values

	H	L
H	204	106
L	294	596

Figure 8.2: *Test Confusion Matrix for Logistic Regression Model 3 (Home Ownership & Employment)*

		Actual Values	
		H	L
Predicted Values	H	48	21
	L	79	152

Figure 8.3: *Accuracy Table for Logistic Regression Model 3 (Home Ownership & Employment)*

	Trained Data	Test Data
Accuracy of model	(204+596)/1200 = 0.666 or 66.6% Accurate	(48+152)/300. = 0.666 or 66.6% Accurate
Accuracy of High App Satisfaction	204/(204+294) = 0.410 or 41.0% Accurate	48/(48+79) = 0.378 or 37.8% Accurate
Accuracy of Low App Satisfaction	596/(596+106) = 0.849 or 84.9% Accurate	152/(152+21) = 0.879 or 87.9% Accurate
Precision	204/(204+106) = 0.658 or 65.8% Precision	48/(48+21) = 0.696 or 69.6% Precision
Misclassification	(294+106)/1200 = 0.333 or 33.3% Error	(79+21)/300 = 0.333 or 33.3% Error

Table 1: *Summary Table for logistic regression on all variables*

Coefficients	Estimate	Standard Error	Z Value	Pr(>|z|)	
Intercept	9.184e-01	5.796e-01	1.585	0.11307	
C_TotalVoluntarySuperContribution	-1.018e-04	2.209e-04	-0.461	0..64506	
C_Age	1.330e-02	1.156e-02	1.151	0.24983	
C_EducationD	2.261e-02	2.327e-01	0.097	0.92261	
C_EducationH	4.462e-01	2.481e-01	1.799	0.07209	
C_EducationM	-5.510e-02	2.578e-01	-0.214	0.83077	
C_EducationP	3.274e-01	2.601e-01	1.259	0.20820	
C_GenderM	1.325e-01	1.645e-01	0.806	0.42053	
C_GenderO	-6.465e-01	9.765e-01	-0.662	0.50798	
C_MonthlyIncome	-5.074e-05	3.683e-05	-1.378	0.16827	
C_HomeOwnershipStatusY	-6.931e-01	2.189e-01	-3.166	0.00155	**
C_RelationshipStatusO	-4.699e-01	2.305e-01	-2.038	0.04151	*
C_RelationshipStatusS	-3.411e-01	2.342e-01	-1.457	0.14521	
C_NumberofDependants	5.160e-03	3.668e-02	0.141	0.88814	
C_EmploymentTypeP	-1.210e-01	1.785e-01	-0.678	0.49785	
C_SalarySacrificeY	-1.995e-01	1.852e-01	-1.077	0.28140	
C_FinancialLiteracyL	-2.367e+00	3.044e-01	-7.778	7.37e-15	***
C_FinancialLiteracyM	-1.081e+00	2.591e-01	-4.171	3.03e-05	***
C_PersonalFinancialAdviceY	6.057e-01	2.308e-01	2.624	0.00868	**
App_Tenure	-1.358e-02	9.826e-03	-1.382	0.16686	
App_AverageDailySessionTime	1.327e-03	4.486e-03	0.296	0.76733	
App_ShowTutorial	2.153e+00	4.104e-01	5.246	1.55e-07	***

Table 2: *Summary Table for Multiple Linear Regression for voluntary superannuation contribution as a function of all variables*

Coefficients	Estimate	Standard Error	T Value	Pr(>|z|)	
Intercept	5.203e+02	7.528e+01	6.911	7.85e-12	***
C_Age	2.257e+00	1.494e+00	1.511	0.131122	
C_EducationD	1.033e+01	2.892e+01	0.357	0.721092	
C_EducationH	-3.239e+01	3.139e+01	-1.032	0.302357	
C_EducationM	-2.974e+01	3.196e+01	-0.931	0.352280	
C_EducationP	5.327e+00	3.288e+01	0.162	0.871350	
C_GenderM	5.588e+00	2.071e+01	0.270	0.787332	
C_GenderO	1.506e+02	1.132e+02	1.331	0.183537	
C_MonthlyIncome	6.659e-02	4.254e-03	15.654	< 2e-16	***
C_HomeOwnershipStatusY	-2.910e+02	2.728e+01	-1.067	0.286322	
C_RelationshipStatusO	-1.773e+02	2.851e+01	-6.220	6.90e-10	***
C_RelationshipStatusS	-2.413e+02	2.909e+01	-8.295	2.94e-16	***
C_NumberofDependants	-1.666e+01	4.576e+00	-3.640	0.000285	***
C_EmploymentTypeP	1.409+01	2.240e+01	0.629	0.529560	
C_SalarySacrificeY	1.867e+02	2.223e+01	8.396	< 2e-16	***
C_FinancialLiteracyL	-4.512e+02	4.315e+01	-10.456	< 2e-16	***
C_FinancialLiteracyM	-1.330e+02	3.510e+01	-3.788	0.000159	***
C_PersonalFinancialAdviceY	-3.918e+01	3.242e+01	-1.209	0.226993	
App_SatisfactionRatingL	2.601e+01	2.829e+01	0.919	0.358035	
App_Tenure	-6.776e-01	1.236e+00	-0.548	0.583677	
App_AverageDailySessionTime	7.892e-01	5.685e-01	1.388	0.165339	
App_ShowTutorial	1.152e+03	4.475e+01	25.739	< 2e-16	***
Multiple R-squared	0.8852				
Adjusted R-squared	0.8831				
F-statistic	432.5 on 21, 1178DF				

Table 3.1: *Summary Table for Simple Linear Regression for voluntary superannuation contribution as a function of Monthly Income*

Coefficients	Estimate	Standard Error	T Value	Pr(>|z|)	
Intercept	-4.151e+02	4.051e+01	-10.25	<2e-16	***
C_MontlyIncome	1.893e-01	4.397e-03	43.06	<2e-16	***
Multiple R-squared	0.6075				
Adjusted R-squared	0.6072				
F-statistic	1854 on 1, 1198 DF				

Table 3.2: *Summary Table for Simple Linear Regression for voluntary superannuation contribution as a function of Financial Literacy*

Coefficients	Estimate	Standard Error	T Value	Pr(>|z|)	
Intercept	2377.18	35.32	67.31	<2e-16	***
C_FinancialLiteracyL	-1977.29	43.19	-45.79	<2e-16	***
C_FinancialLiteracyM	-987.40	50.91	-19.39	<2e-16	***
Multiple R-squared	0.6435				
Adjusted R-squared	0.6429				
F-statistic	1080 on 2, 1197 DF				

Table 3.3: *Summary Table for Multiple Linear Regression for voluntary superannuation contribution as a function of Monthly Income and Financial Literacy*

Coefficients	**Estimate**	**Standard Error**	**T Value**	**Pr(>|z|)**	
Intercept	1.014e+03	6.124e+01	16.56	<2e-16	***
C_MontlyIncome	1.104e-01	4.386e-03	25.16	<2e-16	***
C_FinancialLiteracyL	-1.261e+03	4.506e+01	-27.98	<2e-16	***
C_FinancialLiteracyM	-5.882e+02	4.414e+01	-13.33	<2e-16	***
Multiple R-squared	0.7669				
Adjusted R-squared	0.7663				
F-statistic	1311 on 3, 1196 DF				

Table 4: *Mean Square Error of M*

Variables	**Mean Square Error**
MSE_2 (Monthly Income and Financial Literacy)	247336.2
MSE_3 (Age, Gender, Relationship Status)	490616.3
MSE_all (All variables)	2034277

R-Code

Figure 1

```
data <- read.csv("data.csv")
tvsc <- data$C_TotalVoluntarySuperContribution
income <- data$C_MonthlyIncome
lmIncome <- lm(tvsc ~ income )
options(repr.plot.width = 5, repr.plot.height = 1.5)
par(mar = c(9, 8, 8, 8), cex = 0.65)
plot(income,tvsc, pch = 19, frame = FALSE, ann = FALSE, xlim = c(0,20000), ylim = c(0,4000))
title(xlab = "Monthly Income", ylab = "Total Voluntary Super Contribuition", main = "Monthly Income VS
Voluntary Super Contribuition ")
abline(lmIncome, col = "blue", lwd = 5)
```

Figure 2

```
  data<- na.omit(data)
VoluntaryContribution<-data$C_TotalVoluntarySuperContribution
FinancialLiteracy<-data$C_FinancialLiteracy
FinancialLiteracy = factor (FinancialLiteracy, levels=c("L","M","H"))
boxplot(VoluntaryContribution~FinancialLiteracy,col='cadetblue', main="Financial Literacy vs Voluntary Super
Contribution",cex.main=1)
```

Figure 3

```
  App_satisfaction <- data$App_SatisfactionRating
AST <- data$App_ShowTutorial
APS_AST = table(App_satisfaction, AST)
barplot(APS_AST, las = 2, col = c('blue', 'lightblue'), main = "App satisfaction vs App Show Tutorial",
       ylab = "Number of people", xlab = "App Show Tutorial", ylim = c(0,1500))
par(mar = c(6,4,4,4), cex = 1)
legend_info = c("High", "Low")
legend("topright", legend = legend_info, col = c("blue","lightblue"), lty = 2,
      title = "App satisfaction rating", cex = 0.8, lwd = 6)
```

Figure 4

```
  ASR <- read.csv("data.csv")
table(data$App_SatisfactionRating, data$C_PersonalFinancialAdvice)
mosaicplot(App_SatisfactionRating ~ C_PersonalFinancialAdvice, data = ASR,
          color = c("lightblue", "blue"), main = "App Satisfaction - Personal Financial Advice",
          xlab ="App satisfaction", ylab = "Personal Financial Advice", cex.axis = 1, border = "black")
```

Figure 5.1

```
data <- read.csv("data.csv")
ndata <- nrow(data)
set.seed(57)
```

```
train <- sample(ndata, ndata*0.8)
dataTrain <- data[train, ]
dataTest <- data[-train, ]
lmIncome <- lm(dataTrain$C_TotalVoluntarySuperContribution ~ dataTrain$C_MonthlyIncome, data =
dataTrain)
plot(dataTrain$C_MonthlyIncome, dataTrain$C_TotalVoluntarySuperContribution, xlab = "Monthly Income",
ylab = "Voluntary Super Contribution", main = "Monthly Income vs Voluntary Super Contribution")
abline(lmIncome, col = "blue", lwd = 2)
```

Figure 5.2

```
data <- read.csv("data.csv")
ndata <- nrow(data)
set.seed(57)
train <- sample(ndata, ndata*0.8)
dataTrain <- data[train, ]
dataTest <- data[-train, ]
boxplot(dataTrain$C_TotalVoluntarySuperContribution ~ dataTrain$C_FinancialLiteracy, data = dataTrain, xlab
= "Financial Literacy", ylab = "Voluntary Super Contribution", main = "Financial Literacy vs Voluntary Super
Contribution", col = "blue")
```

Figure 5.3

```
library(readr)
data <- na.omit(data)
ndata <- nrow(data)
set.seed(57)
train <- sample(ndata, ndata*0.8)
dataTrain <- data[train, ]
dataTest <- data[-train, ]
lmIncomeLiteracy <- lm(dataTrain$C_TotalVoluntarySuperContribution ~ dataTrain$C_MonthlyIncome +
dataTrain$C_FinancialLiteracy, data = dataTrain)
dataTrain$C_FinancialLiteracy = factor (dataTrain$C_FinancialLiteracy, levels=c("L","M","H"))
boxplot(dataTrain$C_MonthlyIncome ~ dataTrain$C_FinancialLiteracy, data = dataTrain, xlab = "Monthy
Income", ylab = "Financial Literacy", col='blue')
title( main="Financial Literacy vs Monthly Income")
```

Figure 6 (6.1-6.3)

```
infoo <- read.csv("data.csv")
infoo$Y <- ifelse(infoo$App_SatisfactionRating == "H", 1, 0)
dataTrain <- infoo[1:1200, ]
dataTest <- infoo[1201:1500, ]
logisticSatisfaction <- glm(Y ~ App_ShowTutorial + C_PersonalFinancialAdvice, family = binomial(), data =
infoo)
summary(logisticSatisfaction)
probs <- predict(logisticSatisfaction, newdata = dataTrain, type = "response")
dim(dataTrain)
p <- rep("L",1200)
p[probs > 0.5] <- "H"
```

```
table(p, dataTrain$App_SatisfactionRating)
probsTest <- predict(logisticSatisfaction, newdata = dataTest, type = "response")
dim(dataTest)
pTest <- rep("L", 300)
pTest[probsTest > 0.5] <- "H"
table(pTest, dataTest$App_SatisfactionRating)
```

Figure 7:

```
infoo <- read.csv("data.csv")
infoo$Y <- ifelse(infoo$App_SatisfactionRating == "H", 1, 0)
dataTrain <- infoo[1:1200, ]
dataTest <- infoo[1201:1500, ]
logisticSatisfaction <- glm(Y ~ C_HomeOwnershipStatus + C_EmploymentType, family = binomial(), data =
infoo)
summary(logisticSatisfaction)
probs <- predict(logisticSatisfaction, newdata = dataTrain, type = "response")
dim(dataTrain)
p <- rep("L",1200)
p[probs > 0.5] <- "H"
table(p, dataTrain$App_SatisfactionRating)
probsTest <- predict(logisticSatisfaction, newdata = dataTest, type = "response")
dim(dataTest)
pTest <- rep("L", 300)
pTest[probsTest > 0.5] <- "H"
table(pTest, dataTest$App_SatisfactionRating)
```

Figure 8:

```
infoo <- read.csv("data.csv")
infoo$Y <- ifelse(infoo$App_SatisfactionRating == "H", 1, 0)
dataTrain <- infoo[1:1200, ]
dataTest <- infoo[1201:1500, ]
logisticSatisfaction <- glm(Y ~ C_RelationshipStatus + C_Gender, family = binomial(), data = infoo)
summary(logisticSatisfaction)
probs <- predict(logisticSatisfaction, newdata = dataTrain, type = "response")
dim(dataTrain)
p <- rep("L",1200)
p[probs > 0.5] <- "H"
table(p, dataTrain$App_SatisfactionRating)
probsTest <- predict(logisticSatisfaction, newdata = dataTest, type = "response")
dim(dataTest)
pTest <- rep("L", 300)
pTest[probsTest > 0.5] <- "H"
table(pTest, dataTest$App_SatisfactionRating)
```

Table 1:

```
data$Satisfaction <- ifelse(data$App_SatisfactionRating == "H",1,0)
ndata <- nrow(data)
set.seed(57)
```

```
train <- sample(ndata,ndata*0.8)
SatisfactionTrain <- data[train,]
SatisfactionTest <- data[-train,]
dim(SatisfactionTrain)
dim(SatisfactionTest)
logisticStay <- glm(Satisfaction ~ C_TotalVoluntarySuperContribution + C_Age + C_Education + C_Gender +
C_MonthlyIncome + C_HomeOwnershipStatus + C_RelationshipStatus + C_NumberofDependants +
C_EmploymentType + C_SalarySacrifice + C_FinancialLiteracy + C_PersonalFinancialAdvice + App_Tenure +
App_AverageDailySessionTime + App_ShowTutorial, family = binomial(), data=SatisfactionTrain)
summary(logisticStay)
confint(logisticStay)
```

Table 2:

```
library(readr)
data <- na.omit(data)
ndata <- nrow(data)
set.seed(57)
train <- sample(ndata, ndata*0.8)
dataTrain <- data[train, ]
dataTest <- data[-train, ]
lmall <- lm(dataTrain$C_TotalVoluntarySuperContribution ~ . , data = dataTrain)
summary(lmall)
```

Table 3.1:

```
data <- read.csv("data.csv")
ndata <- nrow(data)
set.seed(57)
train <- sample(ndata, ndata*0.8)
dataTrain <- data[train, ]
dataTest <- data[-train, ]
lmIncome <- lm(dataTrain$C_TotalVoluntarySuperContribution ~ dataTrain$C_MonthlyIncome, data =
dataTrain)
summary(lmIncome)
coef(lmIncome)
cor(dataTrain$C_TotalVoluntarySuperContribution, dataTrain$C_MonthlyIncome)
```

Table 3.2:

```
data <- read.csv("data.csv")
ndata <- nrow(data)
set.seed(57)
train <- sample(ndata, ndata*0.8)
dataTrain <- data[train, ]
dataTest <- data[-train, ]
lmLiteracy <- lm(dataTrain$C_TotalVoluntarySuperContribution ~ dataTrain$C_FinancialLiteracy, data =
dataTrain)
summary(lmLiteracy)
```

```
coef(lmLiteracy)
```

Table 3.3:

```
library(readr)
data <- na.omit(data)
ndata <- nrow(data)
set.seed(57)
train <- sample(ndata, ndata*0.8)
dataTrain <- data[train, ]
dataTest <- data[-train, ]
lmIncomeLiteracy <- lm(dataTrain$C_TotalVoluntarySuperContribution ~ dataTrain$C_MonthlyIncome +
dataTrain$C_FinancialLiteracy, data = dataTrain)
summary(lmIncomeLiteracy)
coef(lmIncomeLiteracy)
```

Table 4: Mean Square Error

```
data <- read.csv("data.csv")
ndata <- nrow(data)
set.seed(57)
train <- sample(ndata, ndata*0.8)
dataTrain <- data[train, ]
dataTest <- data[-train, ]
lmtvsc_2 <- lm(dataTrain$C_TotalVoluntarySuperContribution ~ dataTrain$C_FinancialLiteracy +
dataTrain$C_MonthlyIncome, data = dataTrain)
lmtvsc_3 <- lm(dataTrain$C_TotalVoluntarySuperContribution ~ dataTrain$C_Age +dataTrain$C_Gender +
dataTrain$C_RelationshipStatus, data = dataTrain)
lmall <- lm(C_TotalVoluntarySuperContribution~ ., data = dataTrain)
predtvscTest_2<- predict(lmtvsc_2, newdata = dataTest)
predtvscTest_3<- predict(lmtvsc_3, newdata = dataTest)
predtvscTest_all<- predict(lmall, newdata = dataTest)
MSE_2 <- mean((dataTrain$C_TotalVoluntarySuperContribution - predtvscTest_2)^2)
MSE_3 <- mean((dataTrain$C_TotalVoluntarySuperContribution- predtvscTest_3)^2)
MSE_all <- mean((dataTrain$C_TotalVoluntarySuperContribution- predtvscTest_all)^2)
MSE_2
MSE_3
MSE_all
```