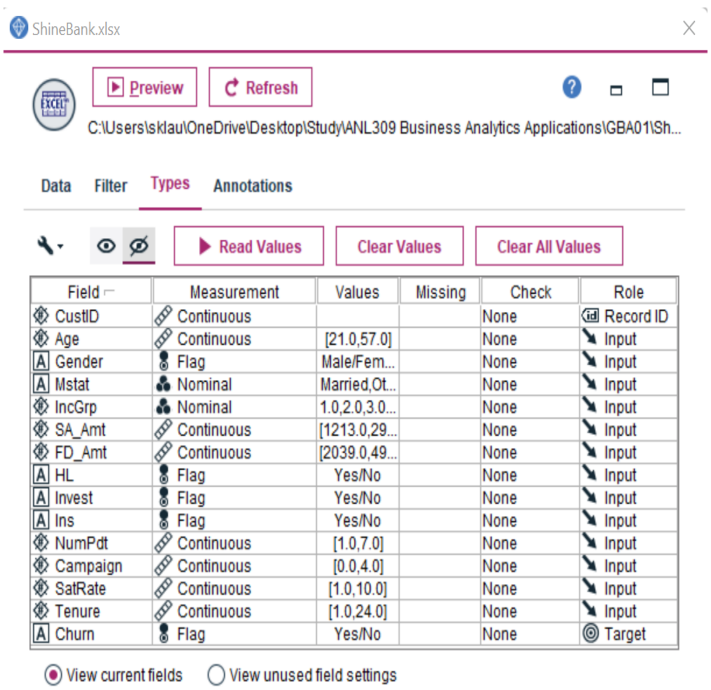
Question 1

a) Data Preparation

The original dataset (Figure 1.1) contains variables of Gender, Mstat, SA_Amt and FD_Amt that are required to be split into separate variables

Figure 1.1

Data type of the original dataset



Field	Measurement	Values	Missing	Check	Role
CustID	Continuous			None	Record ID
Age	Continuous	[21.0,57.0]		None	Input
Gender	Flag	Male/Fem...		None	Input
Mstat	Nominal	Married,Ot...		None	Input
IncGrp	Nominal	1.0,2.0,3.0...		None	Input
SA_Amt	Continuous	[1213.0,29...		None	Input
FD_Amt	Continuous	[2039.0,49...		None	Input
HL	Flag	Yes/No		None	Input
Invest	Flag	Yes/No		None	Input
Ins	Flag	Yes/No		None	Input
NumPdt	Continuous	[1.0,7.0]		None	Input
Campaign	Continuous	[0.0,4.0]		None	Input
SatRate	Continuous	[1.0,10.0]		None	Input
Tenure	Continuous	[1.0,24.0]		None	Input
Churn	Flag	Yes/No		None	Target

The Derive node was used to create the new variables under Gender (Figure 1.2), Mstat (Figure 1.3), SA_Amt (Figure 1.4) and FD_Amt (Figure 1.5)

Figure 1.2

Creation of new variables for Gender

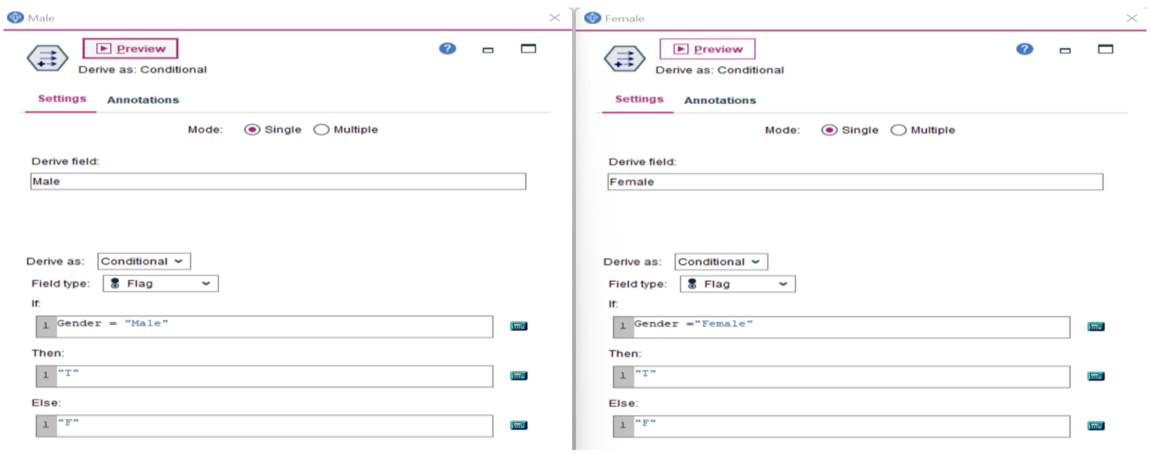


Figure 1.3

Creation of new variables for Mstat

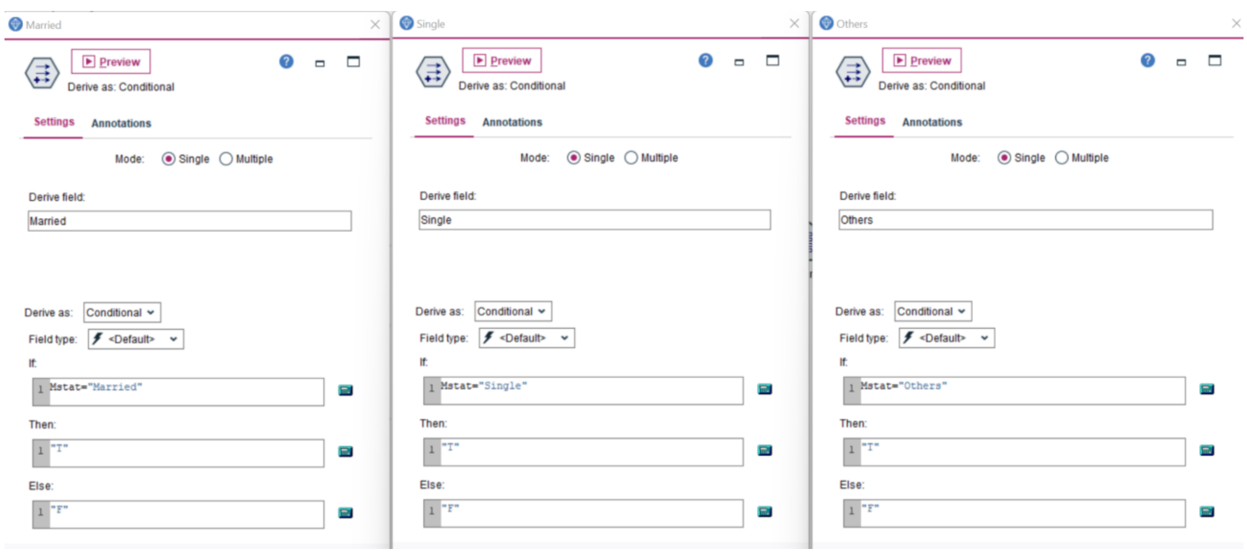


Figure 1.4

Creation of new variables for SA_Amt

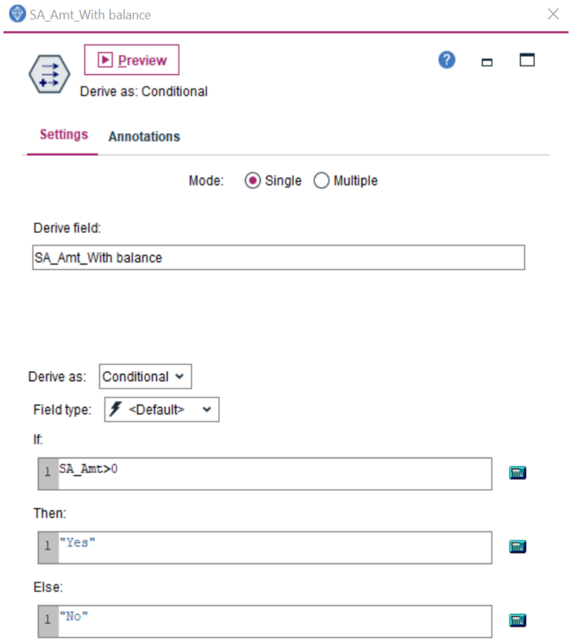
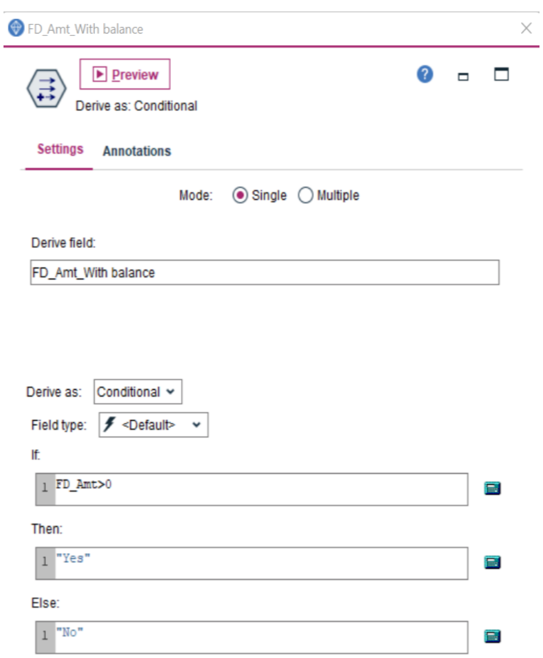


Figure 1.5

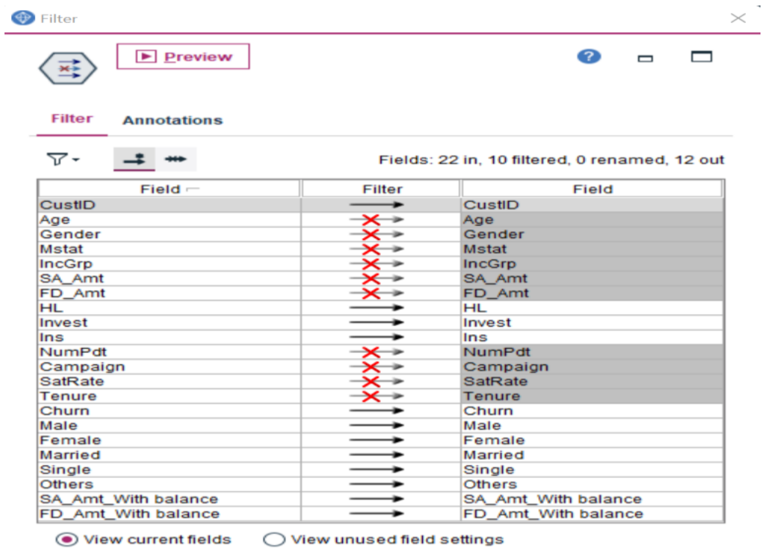
Creation of new variables for FD_Amt



As new variables were created, the Filter node was used to remove the “replaced” variables (Figure 1.6)

Figure 1.6

Filter node used to remove “replaced” variables

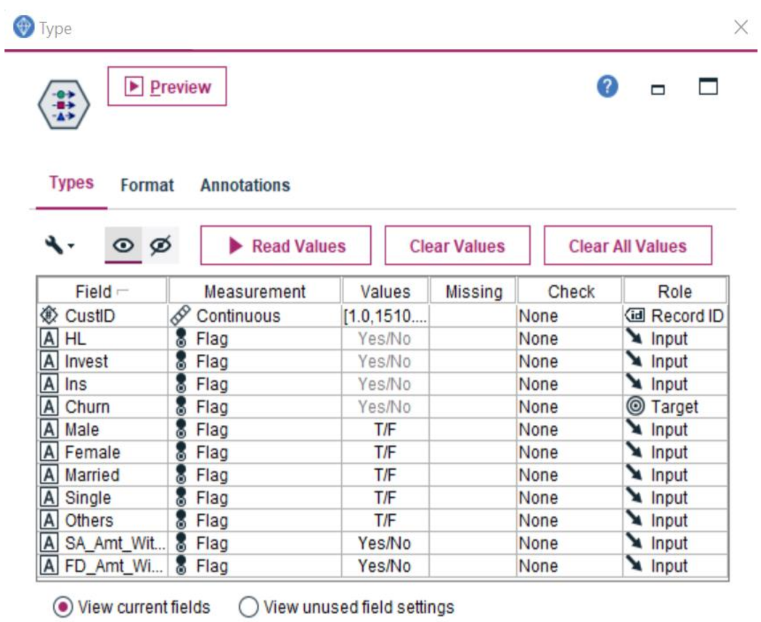


Field	Filter	Field
CustID	→	CustID
Age	✕	Age
Gender	✕	Gender
Mstat	✕	Mstat
IncGrp	✕	IncGrp
SA_Amt	✕	SA_Amt
FD_Amt	✕	FD_Amt
HL	→	HL
Invest	→	Invest
Ins	→	Ins
NumPdt	✕	NumPdt
Campaign	✕	Campaign
SatRate	✕	SatRate
Tenure	✕	Tenure
Churn	→	Churn
Male	→	Male
Female	→	Female
Married	→	Married
Single	→	Single
Others	→	Others
SA_Amt_With balance	→	SA_Amt_With balance
FD_Amt_With balance	→	FD_Amt_With balance

The data type of the newly created variables, along with the unchanged variables for product holdings are shown in Figure 1.7

Figure 1.7

Data type after creation of new variables



Field	Measurement	Values	Missing	Check	Role
CustID	Continuous	[1.0,1510....		None	Record ID
HL	Flag	Yes/No		None	Input
Invest	Flag	Yes/No		None	Input
Ins	Flag	Yes/No		None	Input
Churn	Flag	Yes/No		None	Target
Male	Flag	T/F		None	Input
Female	Flag	T/F		None	Input
Married	Flag	T/F		None	Input
Single	Flag	T/F		None	Input
Others	Flag	T/F		None	Input
SA_Amt_Wit...	Flag	Yes/No		None	Input
FD_Amt_Wi...	Flag	Yes/No		None	Input

b) Using the Apriori node, with an Antecedent Support of 10% and Confidence level of 60% (Figure 1.8), a total of 6 rules was generated (Figure 1.9):

Figure 1.8

Model settings Antecedent Support of 10% and Confidence level of 60%

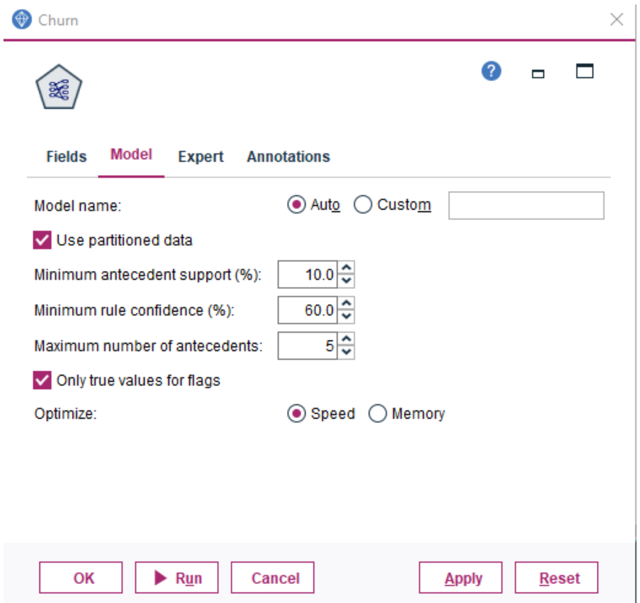
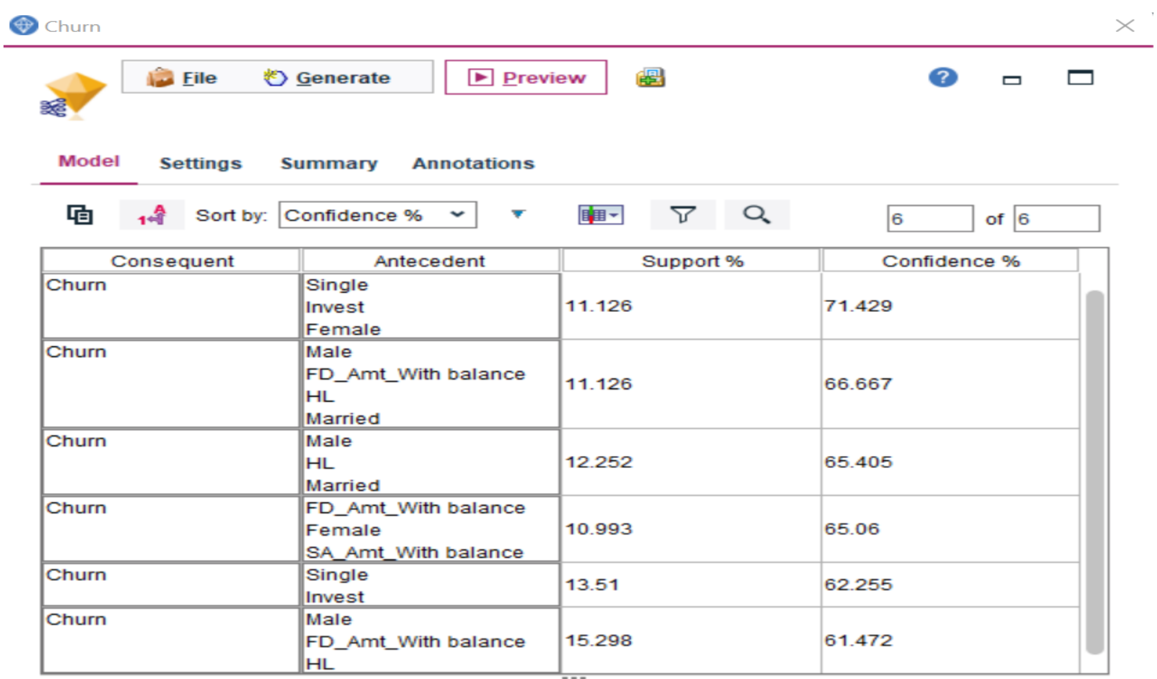


Figure 1.9

Rules generated using the Apriori node



Consequent	Antecedent	Support %	Confidence %
Churn	Single Invest Female	11.126	71.429
Churn	Male FD_Amt_With balance HL Married	11.126	66.667
Churn	Male HL Married	12.252	65.405
Churn	FD_Amt_With balance Female SA_Amt_With balance	10.993	65.06
Churn	Single Invest	13.51	62.255
Churn	Male FD_Amt_With balance HL	15.298	61.472

c) Based on the rules generated by the Apriori node (rules are numbered from 1 to 6, from top to bottom as shown in Figure 1.9), the following rule set can be observed:

Subset rules

1) Rule 1 is a subset of Rule 5
2) Rule 2 is a subset of Rule 3
3) Rule 2 is a subset of Rule 6

Overlapping rules

1) Rule 1 and Rule 4 are overlapping rules
2) Rule 2 and Rule 4 are overlapping rules

Rule sets

Table 1

Customer Group	Rule Set	Rules
1	1	1 and 5
2	2	2, 3 and 6
3	3	4

d) Customer Group 1 contains customers who are single and make investments. There is a 62.255% confidence level that this group of customers will churn on their credit cards. The confidence level increases to 71.429% for female customers under this customer group. To reduce the churn rate, ShineBank may consider increasing the rewards that are linked to using the credit cards. This will incentivize customers who are looking for value when it comes to spending.

Customer Group 2 contains customers who are male and service their home loans with the bank. The confidence level that this group of customers will churn on their credit cards is between 61.472% and 66.667%, depending on whether the customer is married, holds a fixed deposit account with the bank or both. ShineBank can consider packaging credit card usage with fixed deposit account and home loans. Customers who fulfil the criteria will be rewarded with a monthly rebate, up to a certain amount. This will incentivize customers to keep their credit cards.

Customer Group 3 contains customers who are female and maintain both savings and fixed deposit accounts. There is a confidence level of 65.06% that this group of customers will churn on their credit cards. As this group of customers tend to be more conservative and frugal, ShineBank can consider waiving their credit card fees for a number of years. Since no charges will be incurred to maintain the credit cards, customers may be prone to maintain the status quo.

Question 2

a)

Figure 2.1

Data Stream For Clustering

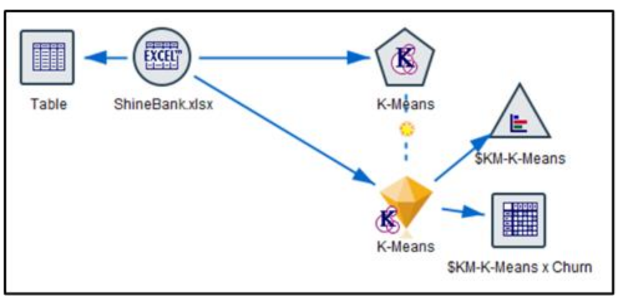


Figure 2.2

Variables used for Clustering

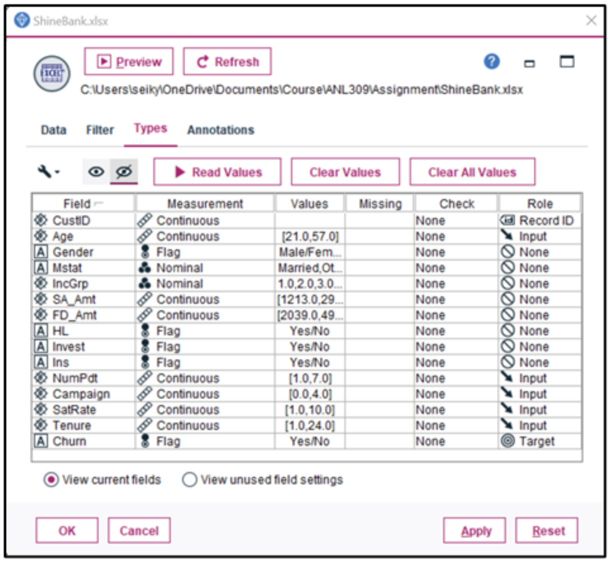


Figure 2.3

K-Means Node Setting

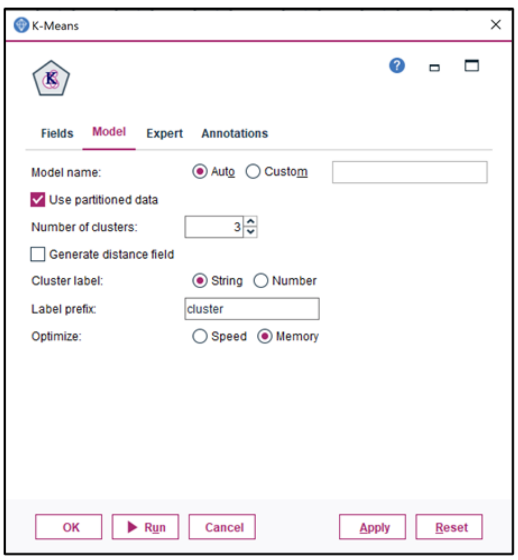


Figure 2.4

K-Means Model Summary

Model Summary

Algorithm	K-Means
Inputs	5
Clusters	3

Cluster Quality

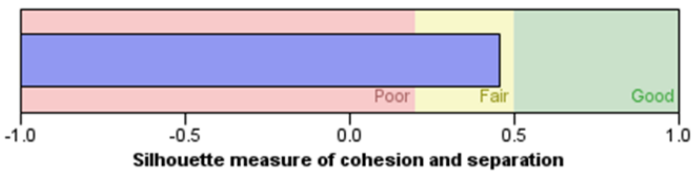
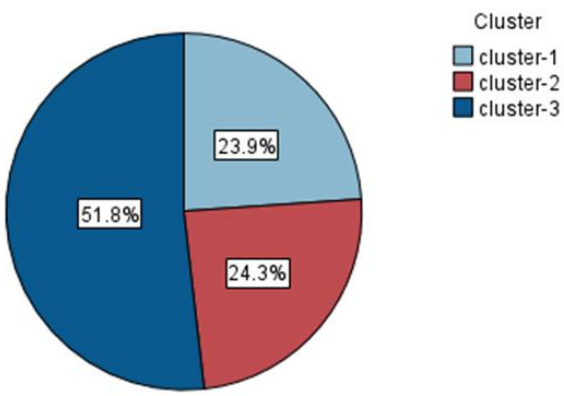


Figure 2.5

K-Mean Cluster Sizes Summary

Cluster Sizes



Size of Smallest Cluster	361 (23.9%)
Size of Largest Cluster	782 (51.8%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	2.17

Figure 2.6

K-Mean Predictor Importance

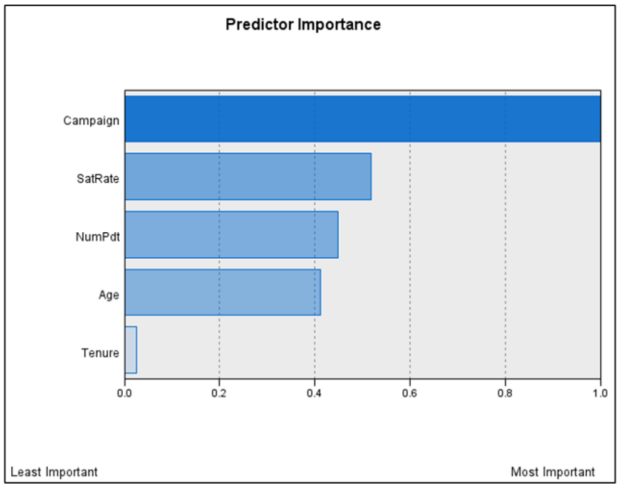


Figure 2.7

Clusters Profile

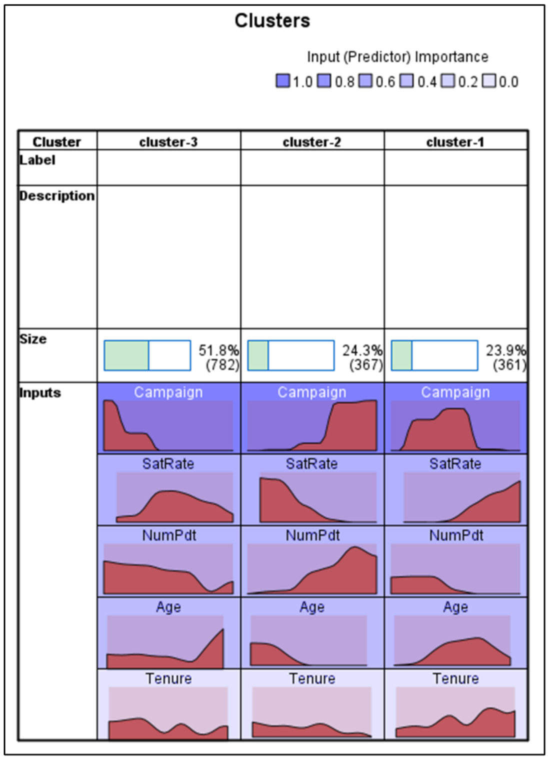
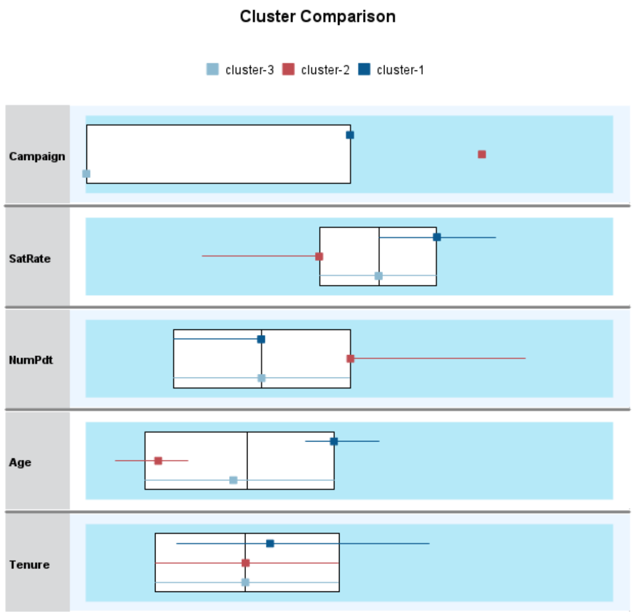


Figure 2.8

Cluster Comparison



b) From Figure 2.4, we can observe that the Cluster Quality is in the Fair region with average silhouette coefficient near 0.5. Furthermore, the size of each cluster is generally big with the smallest cluster at 23.9% and the biggest cluster at the highest incidence 51.8%.

To compare the individual variables characteristic between the cluster from Table 2, and interpretation of clustering results and cluster profiles

Table 2

Cluster Profiles Evaluation

	Cluster-3 : Inactive Moderate Satisfied Working Adult (Inactive)	Cluster-2 : Active UnSatisfied Young Working Adult (Active)	Cluster-1 : Moderate Active Satisfied Older Working Adult (Mod Active)
No of Campaign Customer Participated (Campaign)	Lowest participation relative to Mod Active and Active.	Highest participation relative to Inactive and Active.	Slightly lower participation relative to Active
Customer Satisfaction Rating (SatRate)	Moderate rating relative to Mod Active and Active.	Lower rating relative to Inactive and Mod Active.	Higher rating relative to Inactive and Active.
Number of Product Holdings (NumPdt)	Similar median relative to Mod Active.	Higher numbers relative to Inactive and Mod Active.	Similar median relative to Inactive.
Age (Age)	Slightly lower median relative to Mod Active and Active.	Lowest relative to Mod Active and Inactive	Highest relative to Inactive and Active.
Number of Years with ShineBank (Tenure)	Slightly lower median relative to Mod Active and Active.	Similar median relative to Inactive.	Slightly higher median relative to Inactive and Active

c) The clusters labelled as Inactive Moderate Satisfied Working Adult (Cluster-3), Active Unsatisfied Young Working Adult (Cluster-2) and Moderate Active Satisfied Older Working Adult (Cluster-1).

d) To analyse the incidence of credit card churning in each cluster, we perform a matrix and Distribution graph refer to Figure 2.9 and Figure 2.10.

Figure 2.9

Matrix

Churn

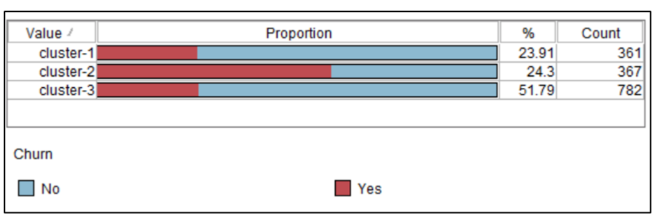
$KM-K-Means		No	Yes	Total
cluster-1	Count	270	91	361
	Row %	74.792	25.208	100
	Column %	26.839	18.056	23.907
	Total %	17.881	6.026	23.907
cluster-2	Count	152	215	367
	Row %	41.417	58.583	100
	Column %	15.109	42.659	24.305
	Total %	10.066	14.238	24.305
cluster-3	Count	584	198	782
	Row %	74.680	25.320	100
	Column %	58.052	39.286	51.788
	Total %	38.675	13.113	51.788
Total	Count	1006	504	1510
	Row %	66.623	33.377	100
	Column %	100	100	100
	Total %	66.623	33.377	100

Cells contain: cross-tabulation of fields (including missing values)

Chi-square = 138.523, df = 2, probability = 0

Figure 2.10

Clusters Distribution



From the figure 2.9, we can observe that Cluster-2 (Active) had the highest churning (58.6%) with relative to Inactive and Mod Active. In this cluster, the customer age is generally younger than other clusters and their satisfaction rating is also relatively lower than other clusters (Inactive, Mod Active). On the other hand, Cluster-1 (Mod Active) had the lowest incidence of Churning (25.2%) relative to Inactive and Active.

Based on the above clustering result, we can make the following recommendation for each cluster profile.

i) Inactive Moderate Satisfied Working Adult (Cluster-3)

As this cluster group characteristic consist almost half of the ShineBank customers and their churn rate is low, ShineBank could focus on strengthening the relationship with them so that their satisfaction level can further increase. ShineBank could provide new products recommendation (new investment trends, insights, children insurance plan, etc) that suit their lifestyle needs and improve their current products holding (more investment returns, upgraded insurance policy, etc). ShineBank could also run simple and straightforward campaign like cashback system that rewards them with staying with the bank. To further improve their customer experience and satisfaction level, ShineBank can introduce a one stop platform that allow customers to easily see all the products holding and perform fast transaction on the same platform.

ii) Active Unsatisfied Young Working Adult (Cluster-2)

As the churn rate for this group (58.6%) is highest relative to Inactive and Mod Active and customers are holding many products, customers are usually demanding and will churn to another bank if the products/ services cannot meet their requirement. To reduce churn rate, ShineBank need to constantly improve their servicing level above market standard so that these customer satisfaction levels will be improved. The service level could be improved by assigning fund manager to smaller group of high product holdings' customers and these fund managers will follow up closely with these customers' needs. ShineBank could also run more aggressive campaign at lower threshold and easily achievable than competitors. By improving the satisfaction level and maintaining the high product holdings, customer churn rate will be reduced.

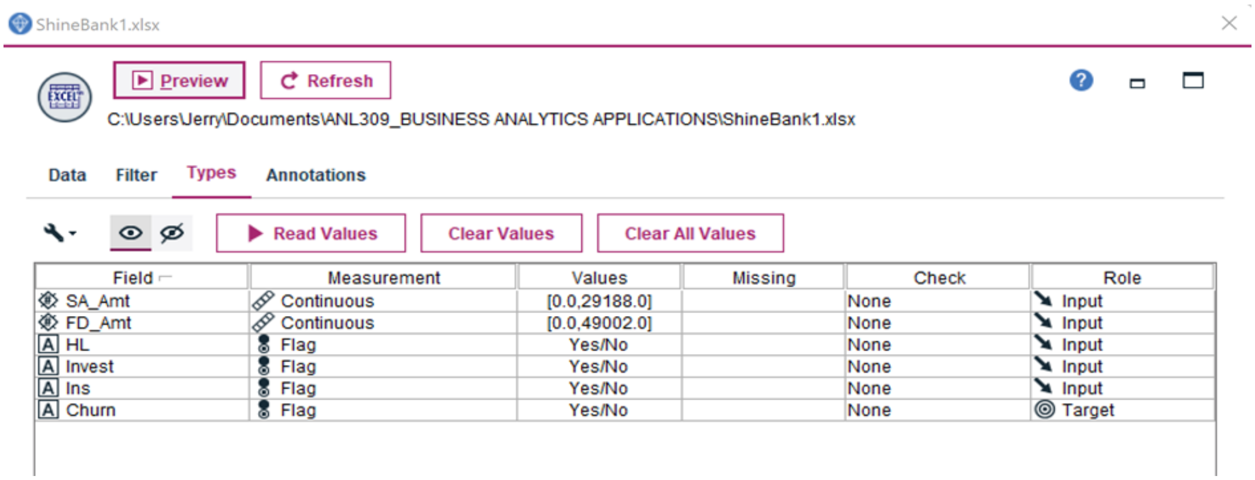
iii) Moderate Active Satisfied Older Working Adult (Cluster-1)

As the characteristic of this group is a more established older age group relative to Inactive and Active, ShineBank could introduce more suitable products that meet their specific needs. For example, insurance policy for more hospitalization coverage. ShineBank could also run campaigns that bundle their products to achieve higher saving interest rates. If customers signed up more products through the campaign, their chances to churn will be reduced.

Question 3

Figure 3.1

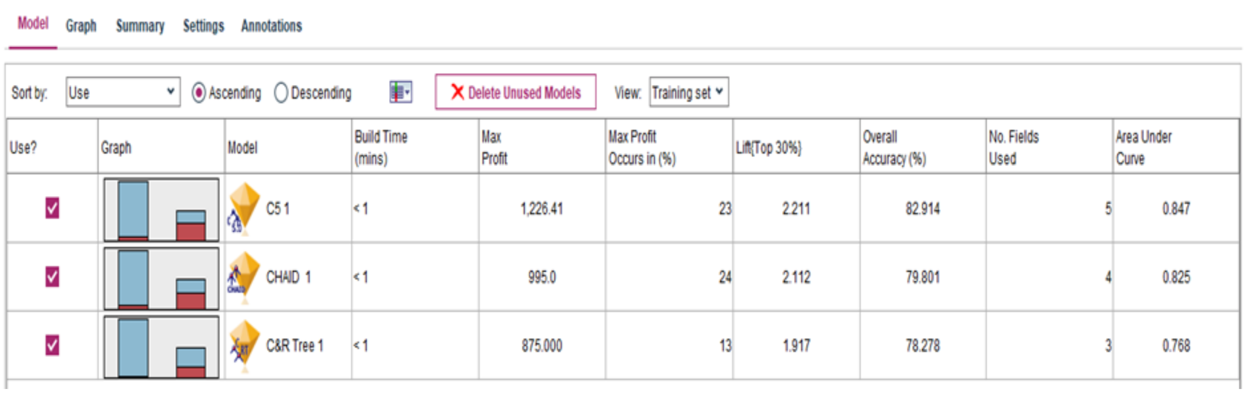
Selected fields' measurement settings to evaluate on a champion decision tree algorithm



Field	Measurement	Values	Missing	Check	Role
SA_Amt	Continuous	[0.0,29188.0]		None	Input
FD_Amt	Continuous	[0.0,49002.0]		None	Input
HL	Flag	Yes/No		None	Input
Invest	Flag	Yes/No		None	Input
Ins	Flag	Yes/No		None	Input
Churn	Flag	Yes/No		None	Target

"Auto Classifier" node is added to the analysis to provide an indication on the "Best" model that can be deployed. The outcome below shows that C5.0 is the "Best" model as it has a relatively higher lift value, overall accuracy and AUC.

Figure 3.2



Use?	Graph	Model	Build Time (mins)	Max Profit	Max Profit Occurs in (%)	Lift{Top 30%}	Overall Accuracy (%)	No. Fields Used	Area Under Curve
✓		C5 1	< 1	1,226.41	23	2.211	82.914	5	0.847
✓		CHAID 1	< 1	995.0	24	2.112	79.801	4	0.825
✓		C&R Tree 1	< 1	875.000	13	1.917	78.278	3	0.768

The Lift chart is examined below on the predictive result of the "churn" category on the "Testing" dataset. At 20% percentile, the lift value of C5.0 (Denoted by the thin red line "$C-Churn") is about 2.5, which is about 2.5 times better than the base-line model. C5.0 has a relatively higher lift value until the 35% percentile, that could suggest C5.0 is a better model in predicting churns compared to the other two models.

Figure 3.3

Lift Chart

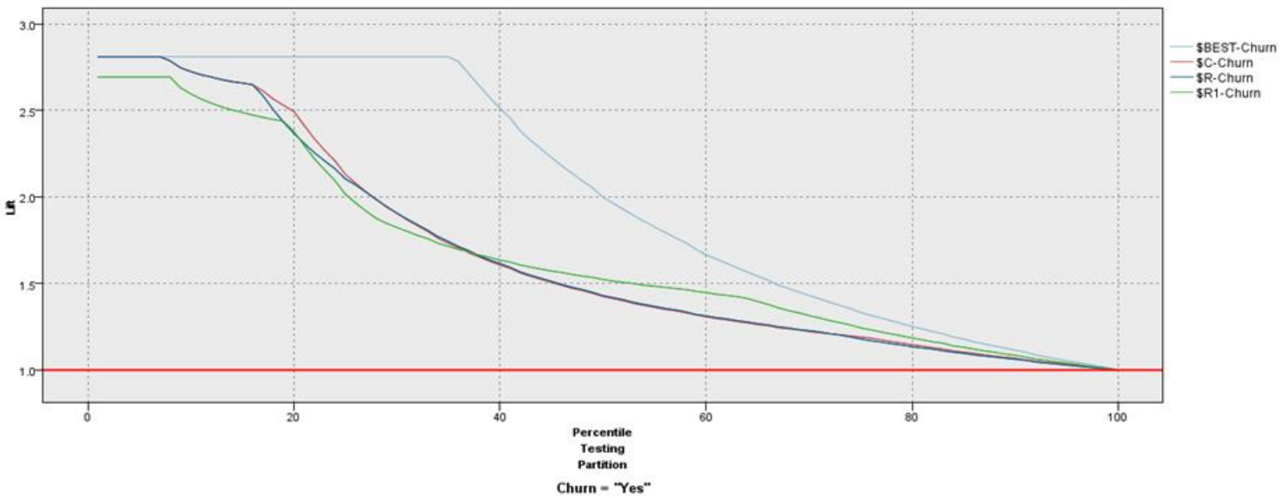


Table 3

Accuracy and hits rates table calculated

	Training Dataset			Testing Dataset		
Metrics	**CHAID**	**CART**	**C5.0**	**CHAID**	**CART**	**C5.0**
Overall Accuracy Rate	77.80%	78.71%	78.88%	78.53%	76.60%	79.81%
Accuracy Rate for Churners	38.42%	49.36%	39.95%	46.85%	54.05%	49.55%
Accuracy Rate for Non-Churners	97.02%	93.04%	97.89%	96.02%	89.05%	96.52%
Hit Rate for Churners	86.29%	77.60%	90.23%	86.67%	73.17%	88.71%
Hit Rate for Non-Churners	76.34%	79.01%	76.95%	76.59%	77.83%	77.60%

C5.0 demonstrates the highest overall accuracy(79.81%), accuracy rate for Non-Churners(96.52%), and hit rates for Churners(88.71%).

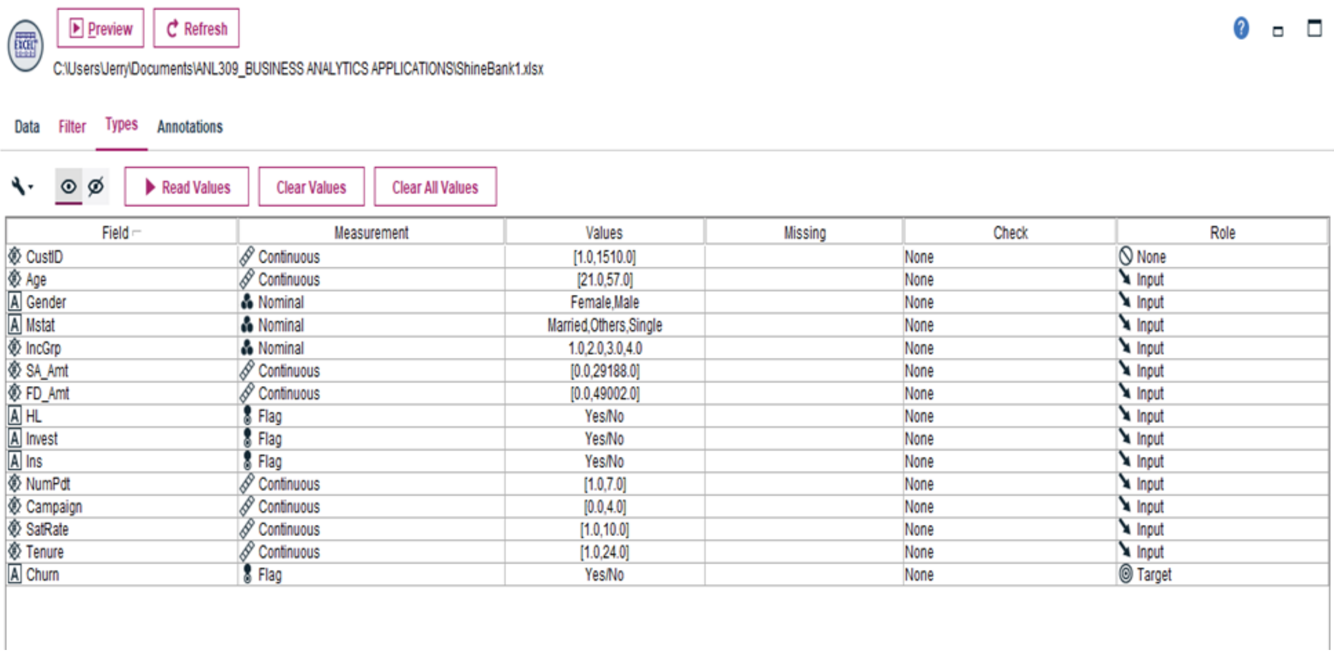
CART has a highest accuracy rate for Churners(54.05%) and hit rate for Non-Churners(77.83%), but has a lowest overall rate, accuracy rate for Non-Churners and Hit Rate for Churners.

If we compare the stability of the model generated under CART and C5.0 for the “training data” against the “testing data”, C5.0 is relatively more consistent on its model performance for most aspects.

Based on the findings above, C5.0 can be deemed as the champion model.

Next, we re-construct the decision tree (C5.0-evaluated as best model) setting all the inputs in the datasets as follows:

Figure 3.4

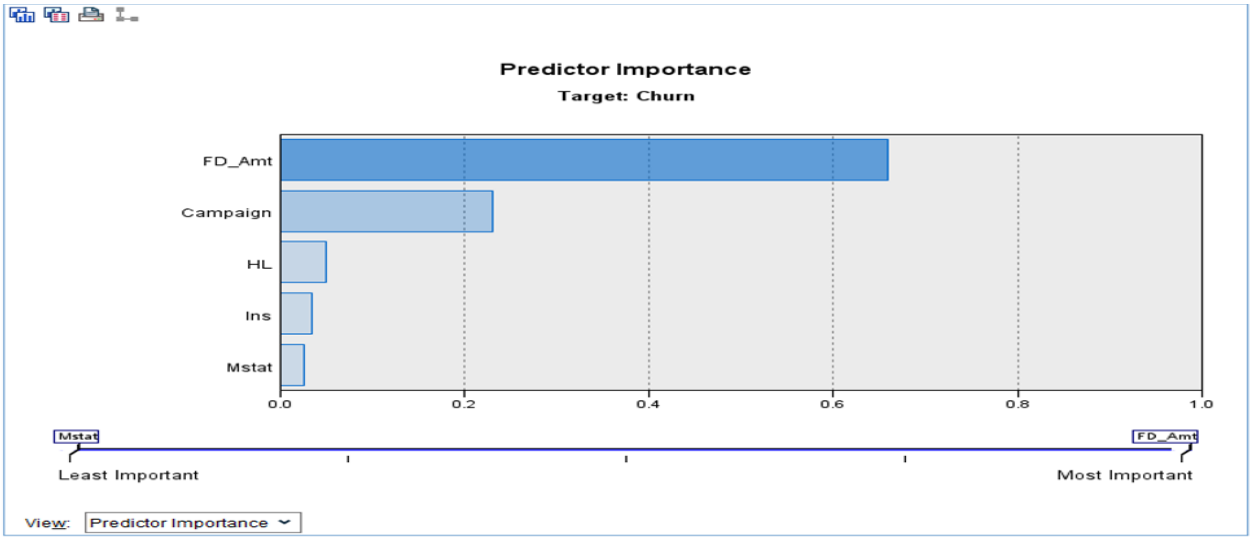


Field	Measurement	Values	Missing	Check	Role
CustID	Continuous	[1.0,1510.0]		None	None
Age	Continuous	[21.0,57.0]		None	Input
Gender	Nominal	Female,Male		None	Input
Mstat	Nominal	Married,Others,Single		None	Input
IncGrp	Nominal	1.0,2.0,3.0,4.0		None	Input
SA_Amt	Continuous	[0.0,29188.0]		None	Input
FD_Amt	Continuous	[0.0,49002.0]		None	Input
HL	Flag	Yes/No		None	Input
Invest	Flag	Yes/No		None	Input
Ins	Flag	Yes/No		None	Input
NumPdt	Continuous	[1.0,7.0]		None	Input
Campaign	Continuous	[0.0,4.0]		None	Input
SatRate	Continuous	[1.0,10.0]		None	Input
Tenure	Continuous	[1.0,24.0]		None	Input
Churn	Flag	Yes/No		None	Target

From the Predictor Importance chart below, “FD_Amt” is at the top of the predictor importance that indicates the highest association to the target (ie the customer will churn or not for the credit card). The next important predictor is the number of campaigns participated by customer in the last 2 years (“Campaign”), followed by any home loan (“HL”), any insurance policy (“Ins”), and marital status (“Mstat”). The rest of the variables may not be that critical in predicting if customer will churn.

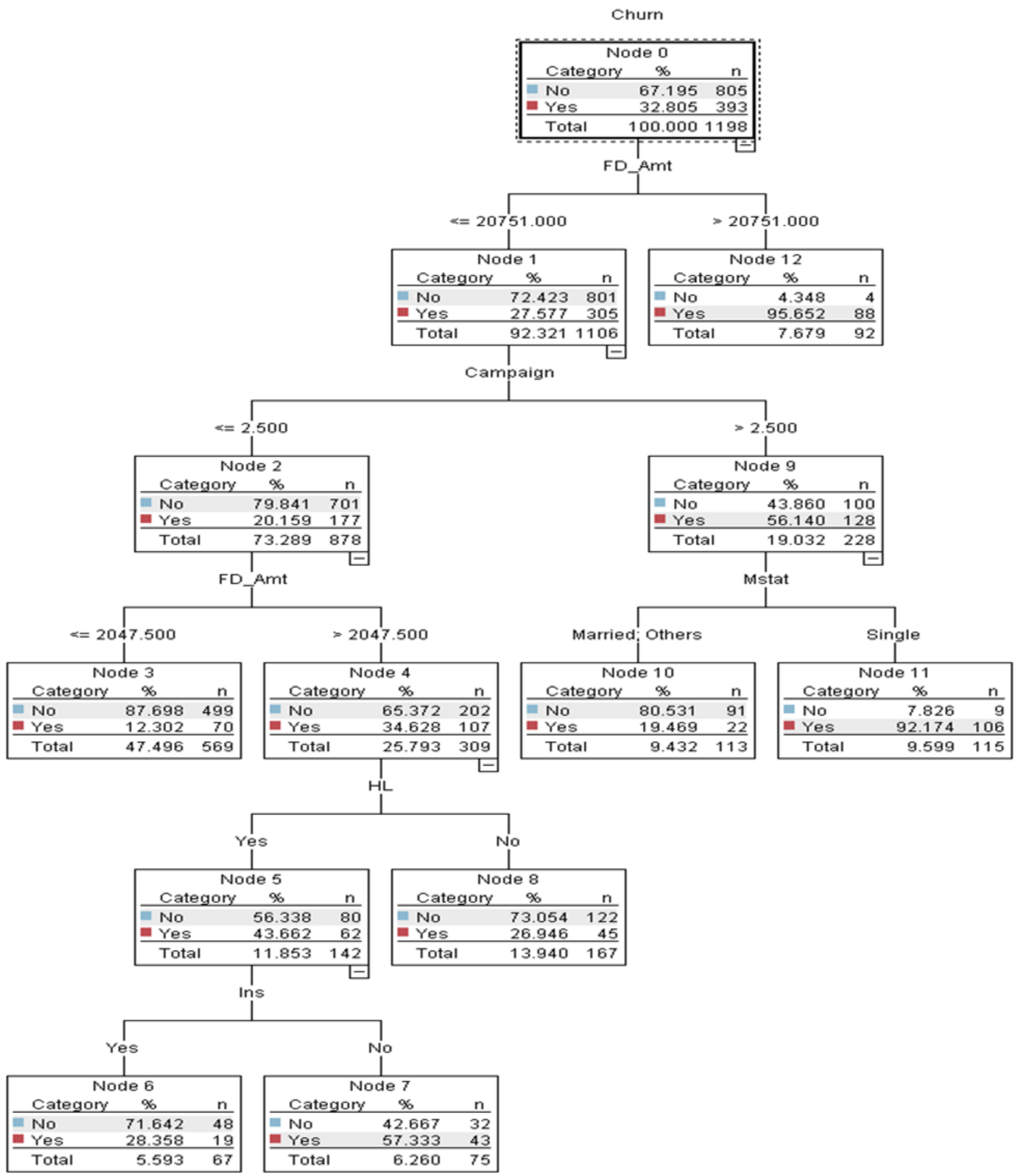
Figure 3.5

Predictor Importance



The decision tree as shown below also indicates that the “FD_Amt” followed by “Campaign” are the more important variables that will determine customer’s churning on his card.

Figure 3.6


Churn
Node 0
Category % n
No 67.195 805
Yes 32.805 393
Total 100.000 1198
FD_Amt
<= 20751.000
> 20751.000
Node 1
Category % n
No 72.423 801
Yes 27.577 305
Total 92.321 1106
Node 12
Category % n
No 4.348 4
Yes 95.652 88
Total 7.679 92
Campaign
<= 2.500
> 2.500
Node 2
Category % n
No 79.841 701
Yes 20.159 177
Total 73.289 878
Node 9
Category % n
No 43.860 100
Yes 56.140 128
Total 19.032 228
FD_Amt
<= 2047.500
> 2047.500
Mstat
Married; Others
Single
Node 3
Category % n
No 87.698 499
Yes 12.302 70
Total 47.496 569
Node 4
Category % n
No 65.372 202
Yes 34.628 107
Total 25.793 309
Node 10
Category % n
No 80.531 91
Yes 19.469 22
Total 9.432 113
Node 11
Category % n
No 7.826 9
Yes 92.174 106
Total 9.599 115
HL
Yes
No
Node 5
Category % n
No 56.338 80
Yes 43.662 62
Total 11.853 142
Node 8
Category % n
No 73.054 122
Yes 26.946 45
Total 13.940 167
Ins
Yes
No
Node 6
Category % n
No 71.642 48
Yes 28.358 19
Total 5.593 67
Node 7
Category % n
No 42.667 32
Yes 57.333 43
Total 6.260 75

The following rules are generated for customers who are likely to churn on his credit card:

i) **Node 12**: Customers who have a FD more than $20,751 has probability of 95.652% churning on their credit cards.

ii) **Node 11**: Customers who have FD less than $20,751, have participated in more than 2 campaigns, and marital status being single, will have a 92.174% probability of churning on their credit cards.

iii) **Node 7**: Customers who have FD more than $2,047.50 but less than $20,751, and participated in less than 3 campaigns, and have home loan but no insurance policy, will have a 57.333% probability of churning on their credit cards.

Recommendations:

ShineBank may need to look into the competitors' fixed deposits and credit cards offerings to understand the underlying churn reasons. Customer could be churning on their credit cards upon completion of their FD tenure. Competitive FD interest rates may be recommended to encourage these customers to extend their FD tenure with ShineBank while offering better credit cards perks.

Currently, ShineBank only offers the same type of credit card to all customers. We can see the higher probability of singles churning their credit cards. There is no differentiation in the perks/ benefits to "Single" customers to "Married/others". ShineBank should consider strengthening its product line offering by understanding different customer needs and wants ie. ShineBank may consider lifestyle benefits credit card targeting at singles.

Question 4

Poster

Credit card churn prediction using Association Analysis, KMeans Clustering, CHAID, CART, C5.0
Choong, H. M. K., Lim, S. J., Lau, S. K., Ong, K. M.

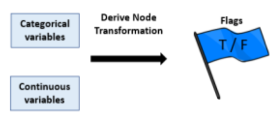
INTRODUCTION

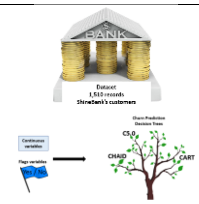
Motivation
- Increase in Credit Card Churning
- Revenue loss
- Expensive to acquire new customers

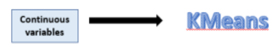
Business Objectives
- To investigate the relationship between input variables with target variable churn
- To understand ShineBank's customer profile
- To predict churn of ShineBank's credit card customers

Data Mining Techniques
- Association analysis perform to discover if customer's demographics variables such as gender and marital status, and product holdings affect churning, using such findings to prevent churning from happening
- KMeans clustering perform for market segmentation according to customer's age, product holdings, campaigns participation, customer satisfaction rating, and customer tenure
- Using CHAID, CART, and C5.0 to predict the best churn model

DATASETS & DATA PREPARATION

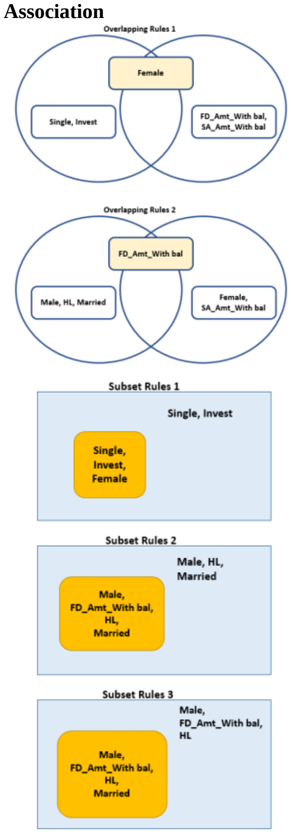




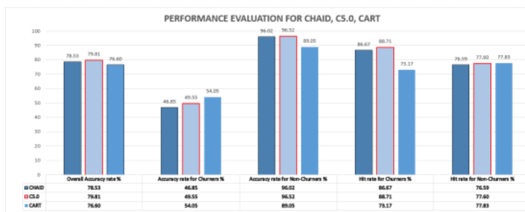


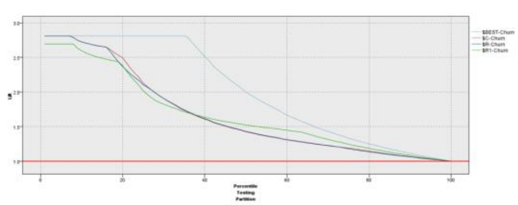
MODELLING AND RESULTS

Association



Predictive Modelling

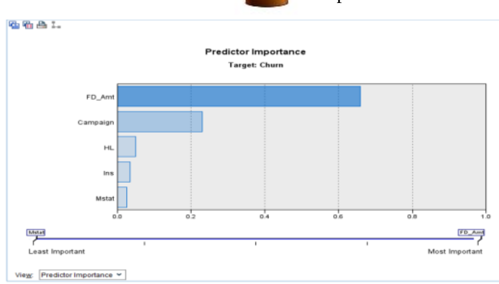




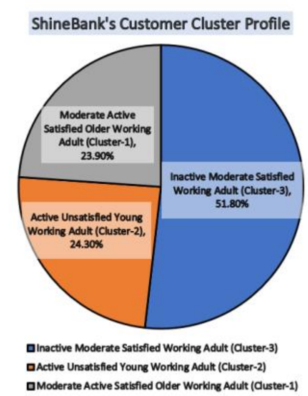
At 20% percentile, C5.0 Lift Value 2.5 times better than base-line churn model

Most important predictor FD_Amt

Least important predictor MStat



KMeans Clustering



Matrix Cluster-2 Highest Churn at 58.6%

RECOMMENDATIONS & CONCLUSION

Customer Group 1
- Increase rewards based upon spending

Customer Group 2
- Bundle credit card usage with fixed deposit and home loans, and rewarding

Customer Group 3
- Waiver of credit card fees for a few years

- Compare competitors' fixed deposits and credit cards offerings to understand the underlying churn reasons
- Give competitive FD interest rates to retain customers
- Offer better credit cards perks
- Introduce lifestyle benefits credit card targeting at singles

Active Unsatisfied Young Working Adult (Prone to Churn)
- Constantly improve customer service level so that customer satisfaction levels will improve
- Assign fund manager to follow up with customers' needs closely
- Run more aggressive campaign

References

Chan, S.P., Tan, S.C., Wang, D., & Ren, J. (2021). *ANL307 Predictive modelling* (study guide). Singapore: Singapore University of Social Sciences. Release V1.7 Build S1.0.5, T1.5.21

Tan, W. C. J. (2021). *ANL309 Business analytics applications* (study guide). Singapore: Singapore University of Social Sciences.

Appendix A

Association Analysis

Figure A1

Association Analysis data mining stream

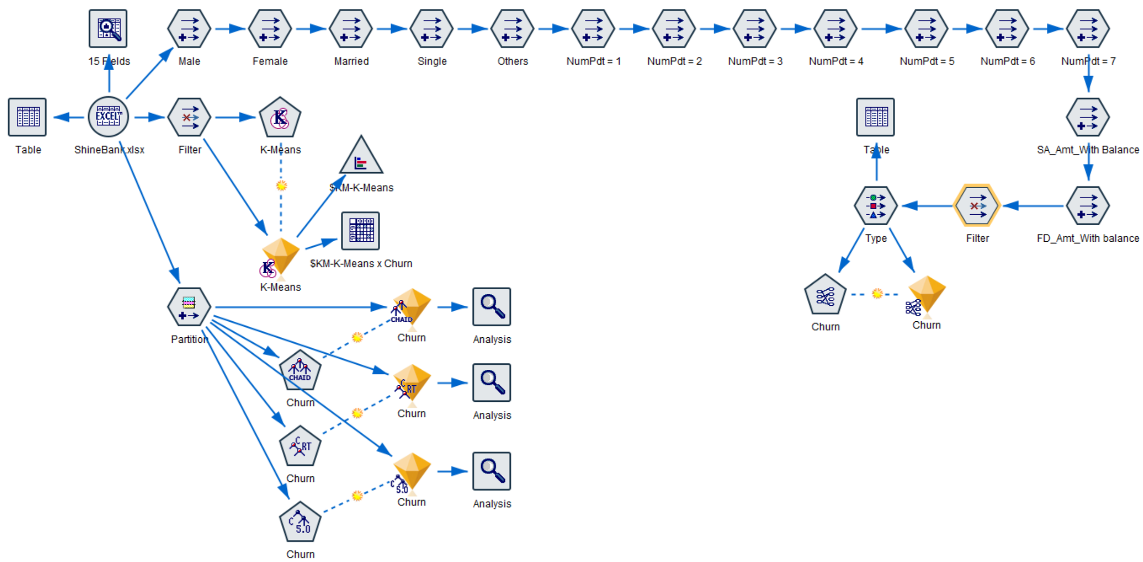


Table A1

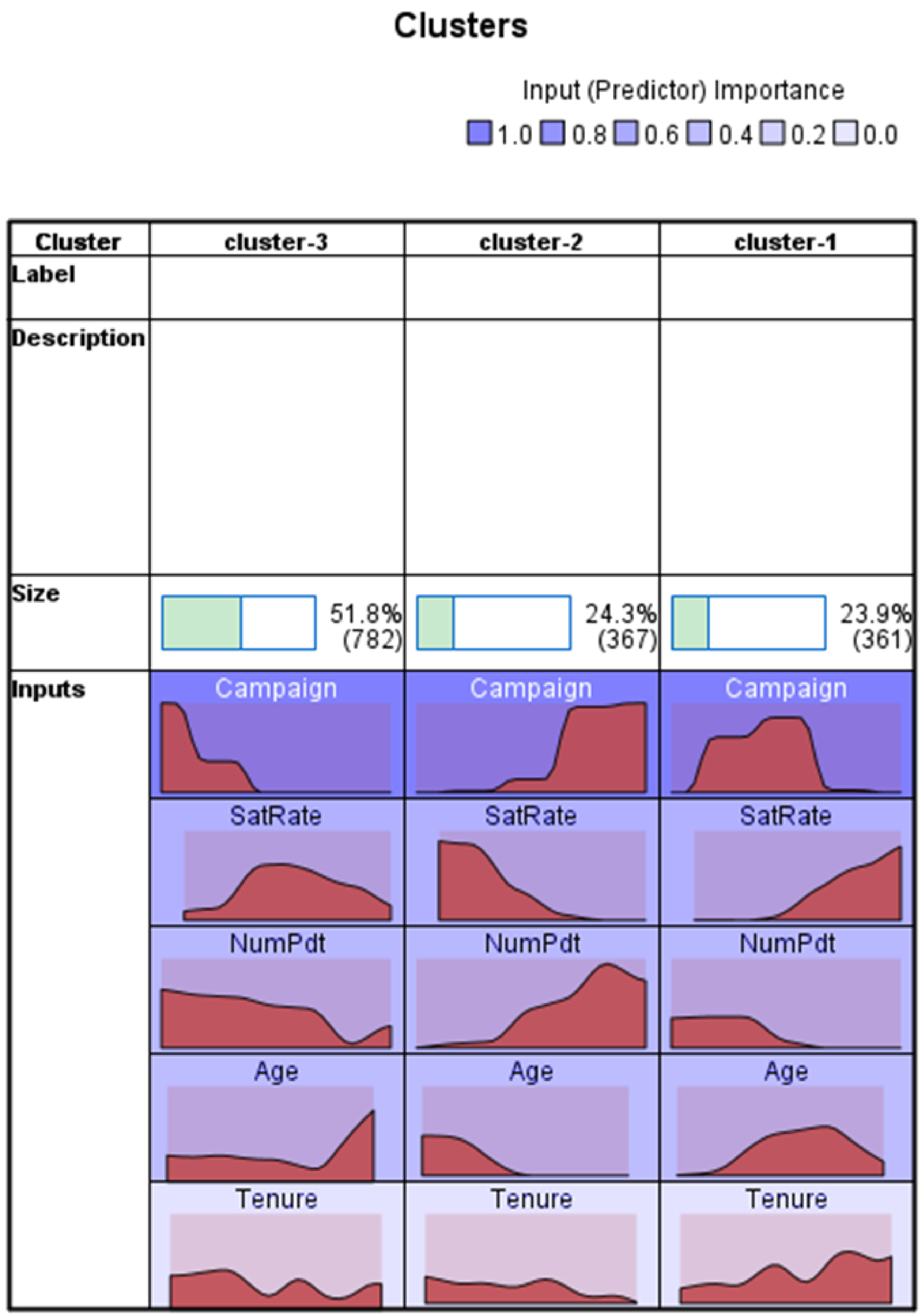
Table node output

Table Annotations

	CustID	HL	Invest	Ins	Churn	Male	Female	Married	Single	Others	SA_Amt_With Balance	FD_Amt_With balance
1	1.000	No	No	Yes	No	T	F	T	F	F	No	No
2	2.000	Yes	Yes	Yes	No	F	T	T	F	F	No	No
3	3.000	Yes	No	Yes	No	T	F	F	T	F	Yes	Yes
4	4.000	Yes	Yes	Yes	No	F	T	T	F	F	Yes	Yes
5	5.000	No	Yes	Yes	No	F	T	T	F	F	No	Yes
6	6.000	No	No	No	No	F	T	F	T	F	No	Yes
7	7.000	Yes	No	Yes	No	F	T	F	T	F	No	No
8	8.000	No	No	Yes	No	T	F	F	F	T	No	No
9	9.000	No	No	No	No	T	F	F	T	F	Yes	No
10	10.000	Yes	No	Yes	No	F	T	F	T	F	No	No
11	11.000	Yes	No	No	Yes	T	F	F	F	T	Yes	Yes
12	12.000	No	No	No	Yes	F	T	F	T	F	Yes	No
13	13.000	Yes	No	No	No	T	F	F	T	F	No	No
14	14.000	No	No	No	No	F	T	F	T	F	Yes	No
15	15.000	No	No	No	No	F	T	F	T	F	Yes	No
16	16.000	No	No	No	No	T	F	F	T	F	No	No
17	17.000	Yes	No	No	Yes	T	F	F	F	T	No	Yes
18	18.000	Yes	No	No	Yes	T	F	F	F	T	Yes	Yes
19	19.000	No	No	No	Yes	F	T	F	T	F	Yes	No
20	20.000	No	No	No	No	F	T	F	T	F	Yes	No
21	21.000	Yes	Yes	No	No	T	F	F	F	T	Yes	Yes
22	22.000	No	No	No	No	F	T	F	T	F	Yes	No
23	23.000	No	No	No	No	F	T	F	T	F	Yes	No
24	24.000	No	No	No	No	F	T	F	T	F	Yes	No
25	25.000	No	Yes	No	No	F	T	F	T	F	No	No
26	26.000	Yes	Yes	Yes	No	T	F	F	F	T	Yes	Yes
27	27.000	Yes	No	No	Yes	T	F	F	F	T	No	Yes
28	28.000	Yes	No	No	Yes	T	F	F	F	T	No	Yes
29	29.000	Yes	No	No	Yes	T	F	F	F	T	No	Yes
30	30.000	Yes	No	No	Yes	T	F	F	F	T	No	Yes
31	31.000	No	No	No	No	F	T	F	T	F	Yes	No
32	32.000	No	No	No	No	F	T	F	T	F	Yes	No
33	33.000	Yes	No	No	Yes	T	F	F	F	T	No	Yes
34	34.000	No	No	Yes	No	F	T	F	T	F	No	No
35	35.000	Yes	No	No	Yes	T	F	F	F	T	No	Yes
36	36.000	Yes	No	No	Yes	T	F	F	F	T	No	Yes
37	37.000	No	No	Yes	Yes	F	T	F	T	F	Yes	No
38	38.000	No	No	Yes	No	F	T	F	T	F	Yes	No
39	39.000	No	No	Yes	No	F	T	F	T	F	Yes	Yes
40	40.000	No	No	No	No	F	T	F	T	F	Yes	No
41	41.000	No	No	No	Yes	F	T	F	T	F	Yes	No
42	42.000	Yes	No	No	Yes	T	F	F	F	T	No	Yes
43	43.000	No	No	No	No	F	T	F	T	F	Yes	No
44	44.000	No	No	No	Yes	F	T	F	T	F	Yes	No
45	45.000	No	No	No	Yes	F	T	F	T	F	Yes	No
46	46.000	No	No	Yes	No	F	T	F	T	F	Yes	No
47	47.000	No	No	Yes	No	F	T	F	T	F	Yes	No
48	48.000	Yes	No	No	Yes	T	F	F	F	T	No	Yes
49	49.000	No	No	Yes	No	F	T	F	T	F	No	No
50	50.000	Yes	No	Yes	No	T	F	F	F	T	No	Yes

Appendix B

Clustering Profiles



Appendix C

Predictive Modelling

Figure C1

Evaluating the 3 models

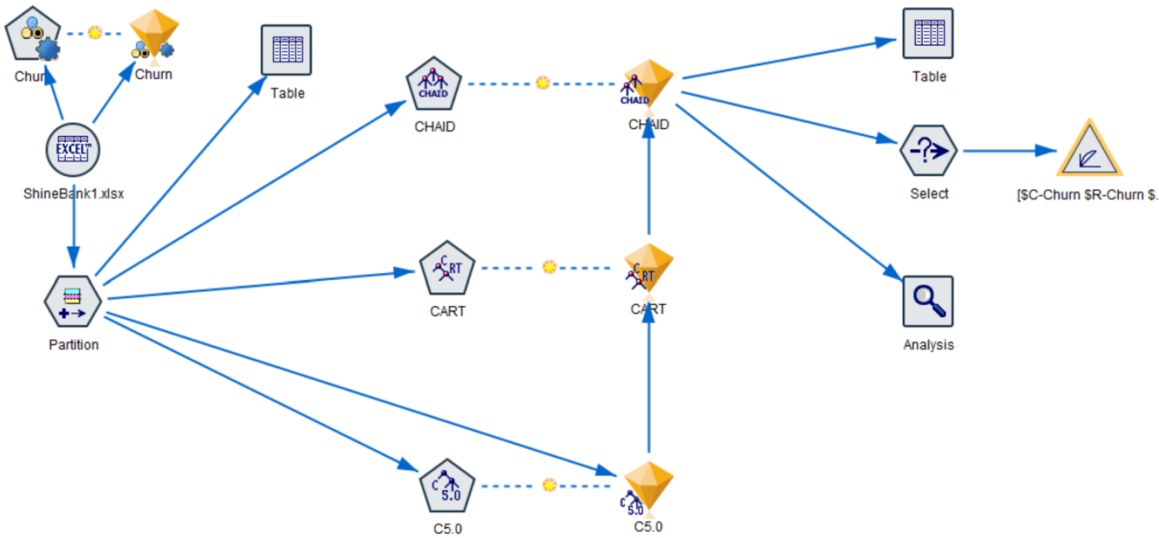


Figure C2

C5.0 Model for all inputs

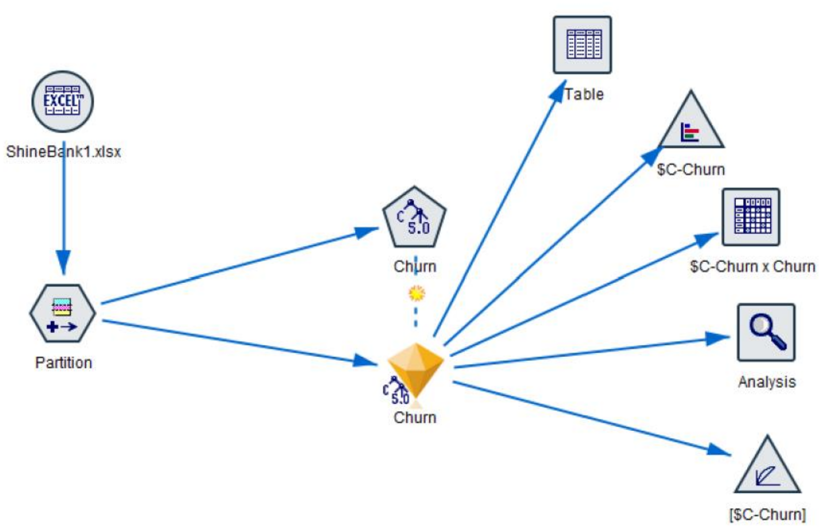


Figure C3

Partition Node

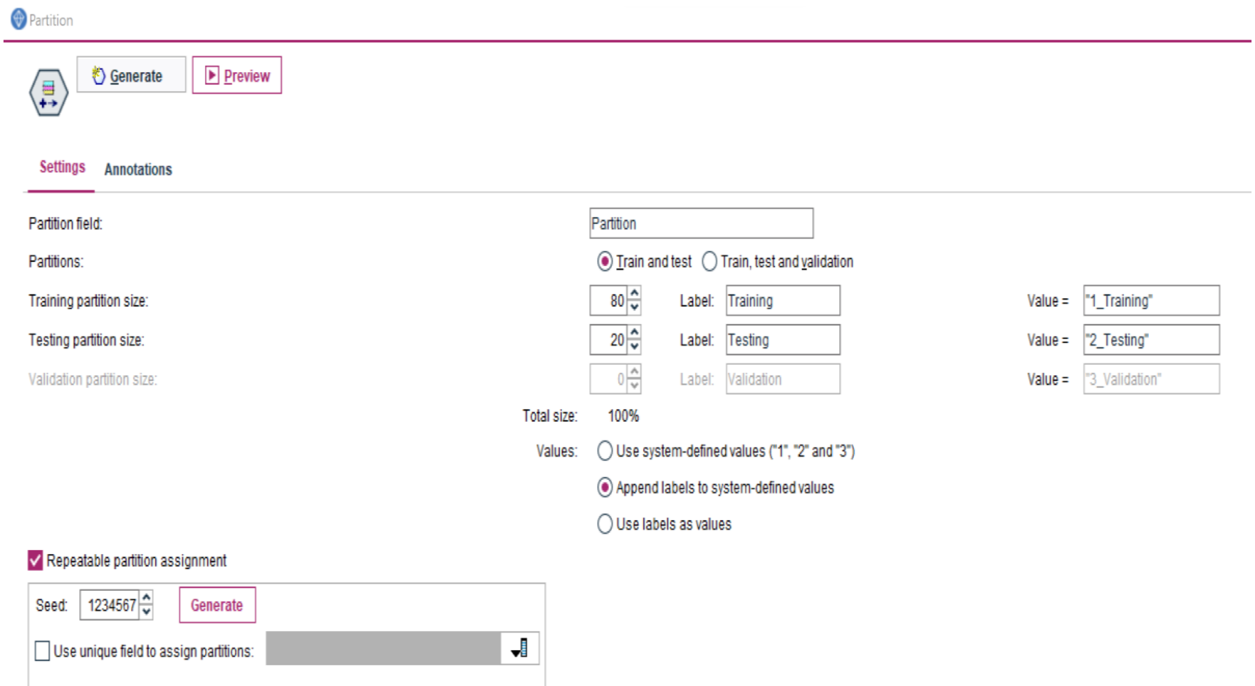


Figure C4

Chaid Setting

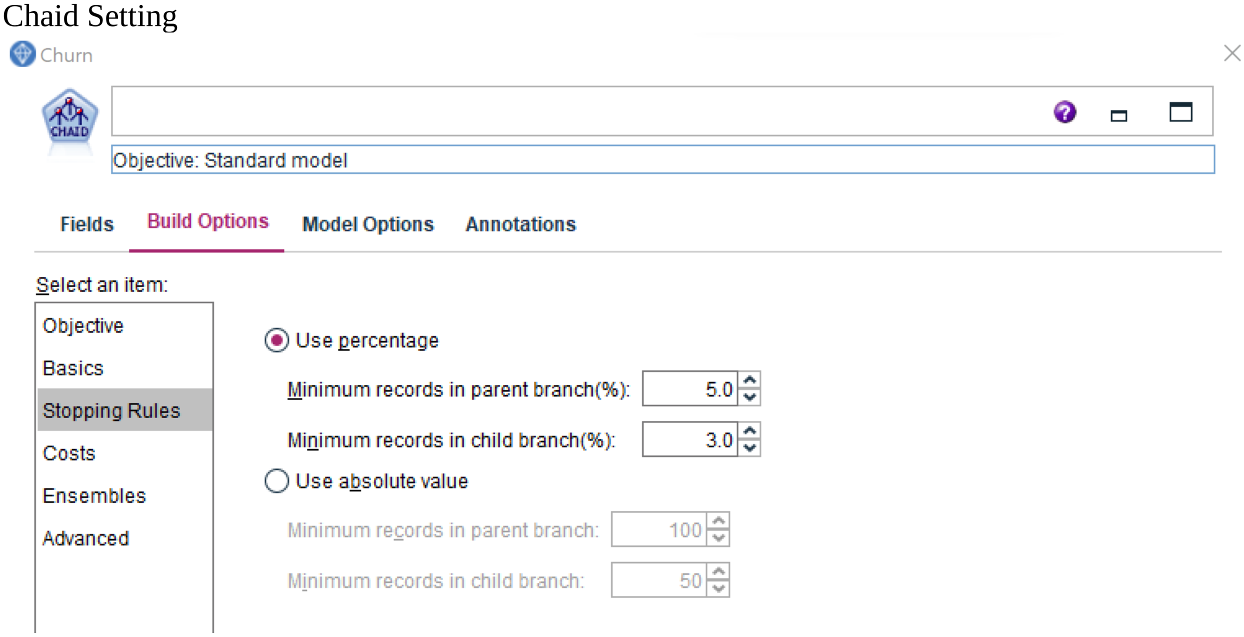


Figure C5

CART Setting

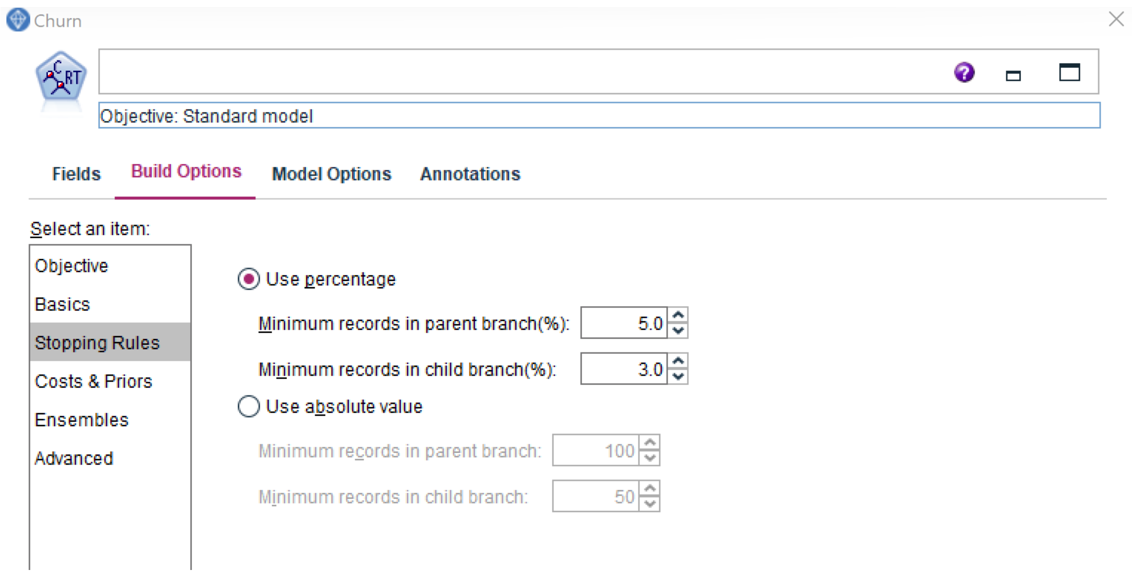


Figure C6

CART Setting

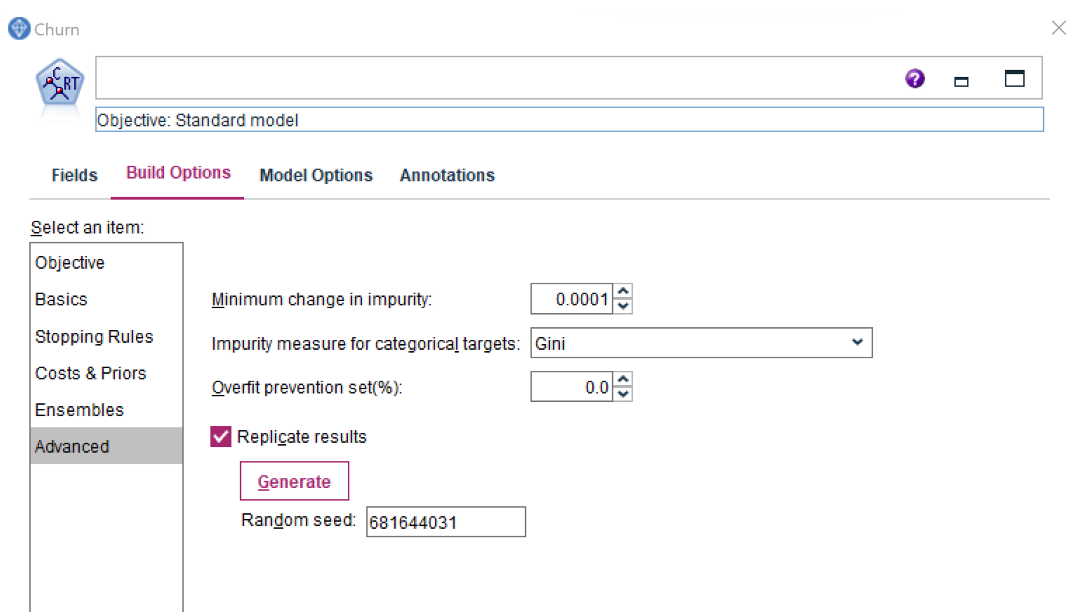


Figure C7

C5.0 Setting