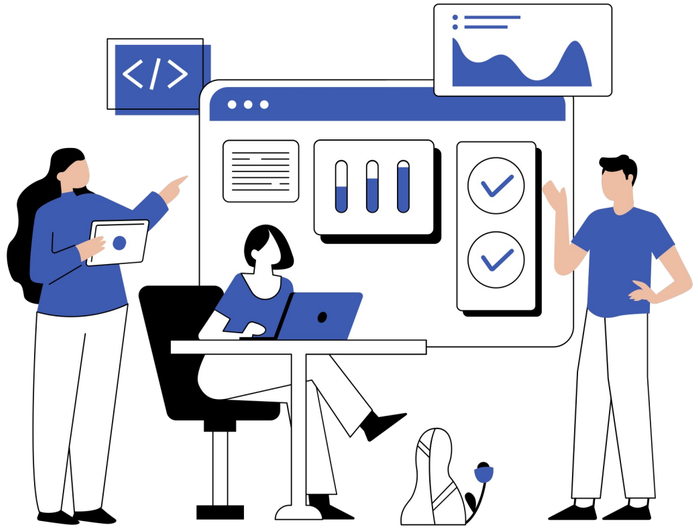
KATION

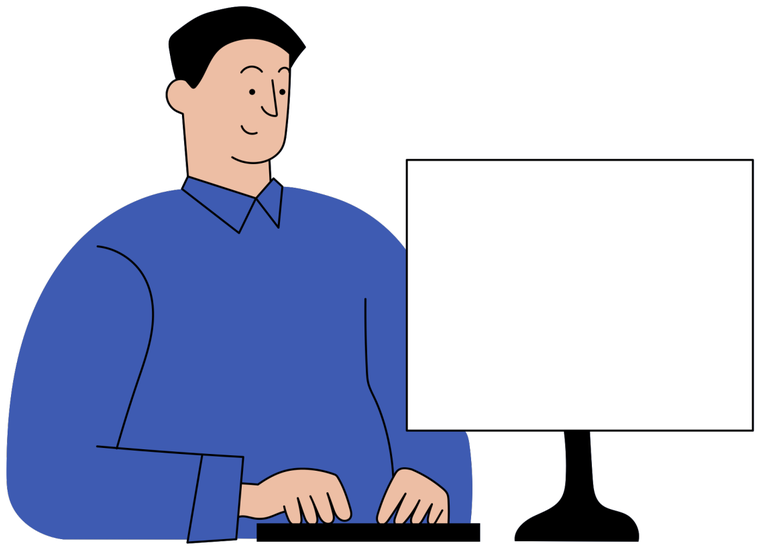
WQD 7005 DATA MINING



Evaluating Telco Campaign Performance And Predicting Campaign Offer Takers

KHAR SHIN YIN (S2145827) LEE SHIN EE (S2132380) OOI SHI YUAN (17088226) SIM LIN ZHENG (S2102170) YU YUEN HERN (S2121801)

Content



Introduction

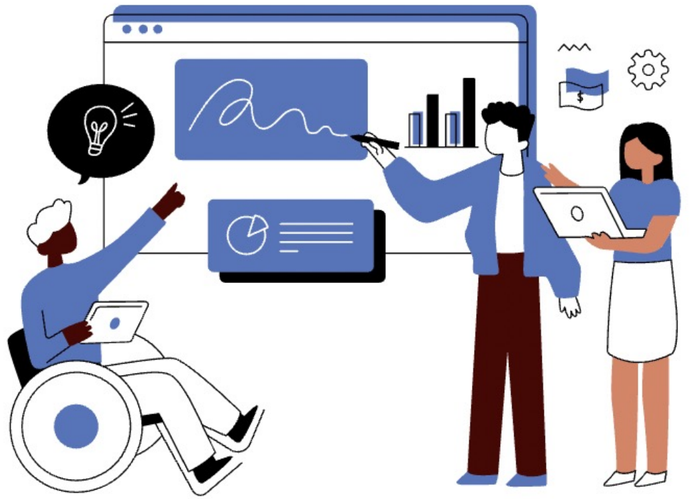
Platform migration in Kation

- An initiative called ‘Right Planning’ is conducted
  - to migrate old rate plans to new rate plans
- Business goal: To standardize the rate plans & enhance customers’ experience and users’ experience.
- Decided to launch a pilot campaign.

Analysis Goal & Objective

To assess pilot campaign performance before implementing to the entire customer base.

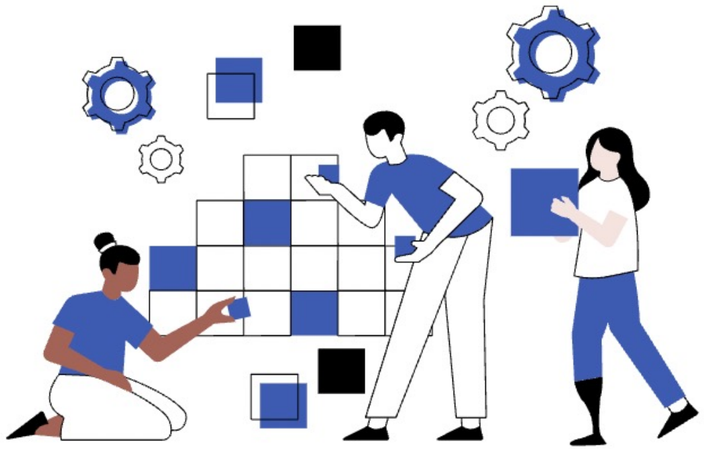
- To evaluate the effectiveness of ‘Right Planning’ pilot campaign.
- To determine campaign takers’ profile.
- To predict campaign takers based on usage and revenue behavior.



Dataset Description

Campaign data in customer level

- Demographic
- Usage before and after campaign
- Revenue before and after campaign
- Status before and after campaign



Demographic & Status

ID
TENURE
AGE
GENDER
NATIONALITY
STATE
STATUS_BEFORE
STATUS_AFTER
OFFER_TAKER
OFFER_TAKE_UP_DT
ACTIVITY_DAYS_AFTER
ACTVIITY_STATUS_AFTER

Usage

DATA_USG_BEFORE
DATA_USG_AFTER
DATA_PURC_BEFORE
DATA_PURC_AFTER
RLD_IND_BEFORE
RLD_IND_AFTER
VOICE_USG_BEFORE
VOICE_USG_AFTER

Revenue

ARPU_BEFORE
ARPU_AFTER
CPA_RVN_BEFORE
CPA_RVN_AFTER
DATA_CHRG_BEFORE
DATA_CHRG_AFTER
RLD_AMT_BEFORE
RLD_AMT_AFTER

Sample

7272 records, 27 features

Basic

Metadata Completed.

Library: DATASET
Data Source: Dataset
Role: Raw

Role	Level	Count
Input	Interval	7
Input	Nominal	20

Advance

Metadata Completed.

Library: DATASET
Data Source: Dataset
Role: Raw

Role	Level	Count
Input	Binary	5
Input	Interval	7
Input	Nominal	5
Rejected	Nominal	8
Rejected	Unary	2

Sample

Reclassification

Name	Role	Level
ACTIVITY_STATUS_AFTER	INPUT	NOMINAL
ACTVIITY_DAYS_AFTER	INPUT	INTERVAL
AGE	INPUT	INTERVAL
ARPU_AFTER	REJECTED	NOMINAL
ARPU_BEFORE	REJECTED	NOMINAL
CPA_RVN_AFTER	REJECTED	NOMINAL
CPA_RVN_BEFORE	REJECTED	NOMINAL
DATA_CHRG_AFTER	REJECTED	NOMINAL
DATA_CHRG_BEFORE	REJECTED	NOMINAL
DATA_PURC_AFTER	INPUT	BINARY
DATA_PURC_BEFORE	INPUT	BINARY
DATA_USG_AFTER	INPUT	INTERVAL
DATA_USG_BEFORE	INPUT	INTERVAL
GENDER	INPUT	NOMINAL
NATIONALITY	REJECTED	UNARY
OFFER_TAKER	INPUT	NOMINAL
OFFER_TAKE_UP_DT	INPUT	BINARY
RLD_AMT_AFTER	REJECTED	NOMINAL
RLD_AMT_BEFORE	REJECTED	NOMINAL
RLD_IND_AFTER	INPUT	BINARY
RLD_IND_BEFORE	INPUT	BINARY
STATE	INPUT	NOMINAL
STATUS_AFTER	INPUT	NOMINAL
STATUS_BEFORE	REJECTED	UNARY
TENURE	INPUT	INTERVAL
VOICE_USG_AFTER	INPUT	INTERVAL
VOICE_USG_BEFORE	INPUT	INTERVAL

Advance

Name	Role	Level
ACTIVITY_STATUS_AFTER	INPUT	NOMINAL
ACTVIITY_DAYS_AFTER	INPUT	INTERVAL
AGE	INPUT	INTERVAL
ARPU_AFTER	INPUT	NOMINAL
ARPU_BEFORE	INPUT	NOMINAL
CPA_RVN_AFTER	INPUT	NOMINAL
CPA_RVN_BEFORE	INPUT	NOMINAL
DATA_CHRG_AFTER	INPUT	NOMINAL
DATA_CHRG_BEFORE	INPUT	NOMINAL
DATA_PURC_AFTER	INPUT	BINARY
DATA_PURC_BEFORE	INPUT	BINARY
DATA_USG_AFTER	INPUT	INTERVAL
DATA_USG_BEFORE	INPUT	INTERVAL
GENDER	INPUT	NOMINAL
NATIONALITY	INPUT	NOMINAL
OFFER_TAKER	TARGET	BINARY
OFFER_TAKE_UP_DT	INPUT	NOMINAL
RLD_AMT_AFTER	INPUT	NOMINAL
RLD_AMT_BEFORE	INPUT	NOMINAL
RLD_IND_AFTER	INPUT	BINARY
RLD_IND_BEFORE	INPUT	BINARY
STATE	INPUT	NOMINAL
STATUS_AFTER	INPUT	NOMINAL
STATUS_BEFORE	INPUT	NOMINAL
TENURE	INPUT	INTERVAL
VOICE_USG_AFTER	INPUT	INTERVAL
VOICE_USG_BEFORE	INPUT	INTERVAL

Manual

Explore

Univariate Analysis

STATE
GENDER

Inconsistent

AGE

Intentional

ARPU_BEFORE
ARPU_AFTER
CPA_RVN_BEFORE
CPA_RVN_AFTER
DATA_CHRG_BEFORE
DATA_CHRG_AFTER
RLD_AMT_BEFORE
RLD_AMT_AFTER

Noisy & Incomplete

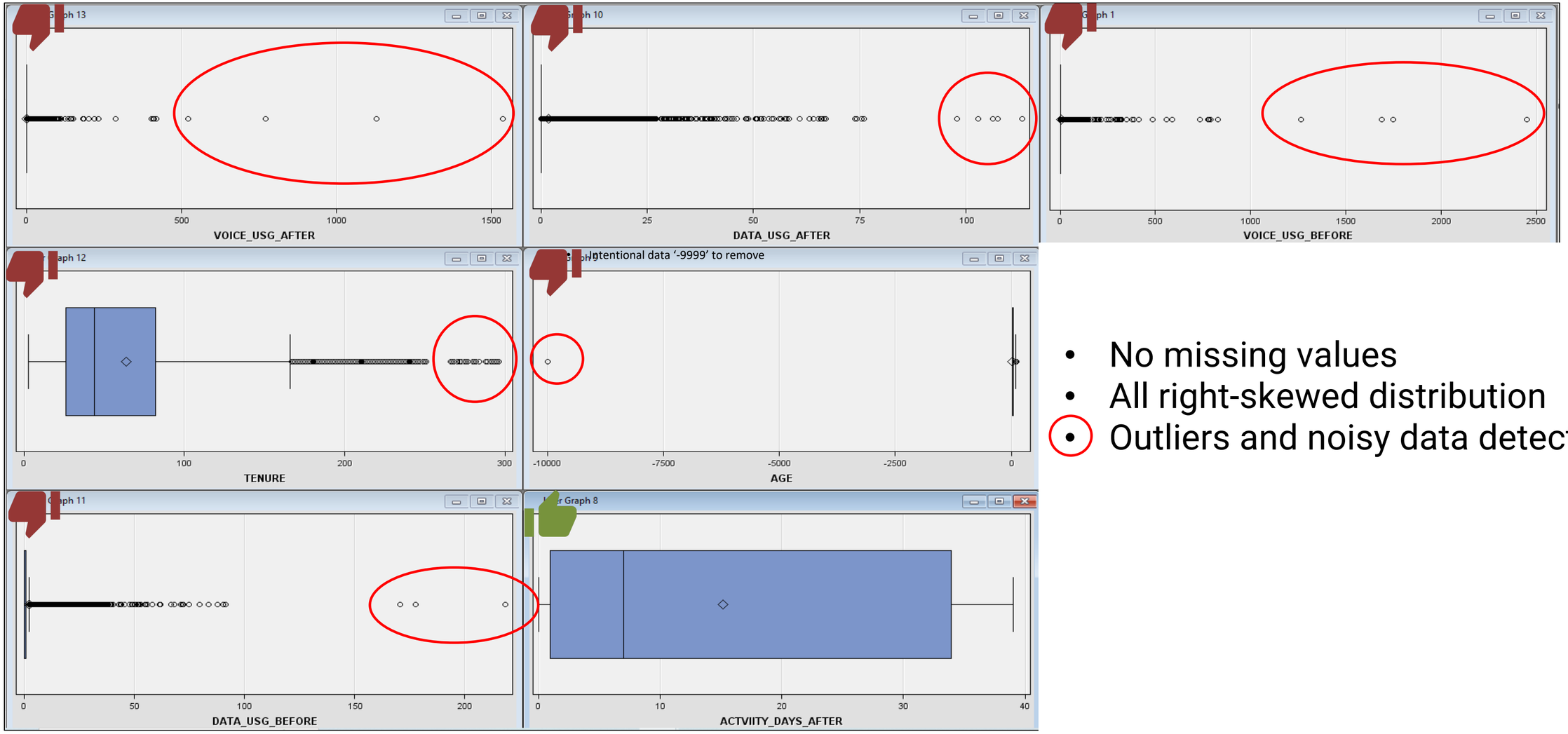
OFFER_TAKE_UP_DT

Incomplete

- There are 4 issues found: Intentional, Noisy, Incomplete & Inconsistent data
- Perform data cleaning during Modify stage.

Explore

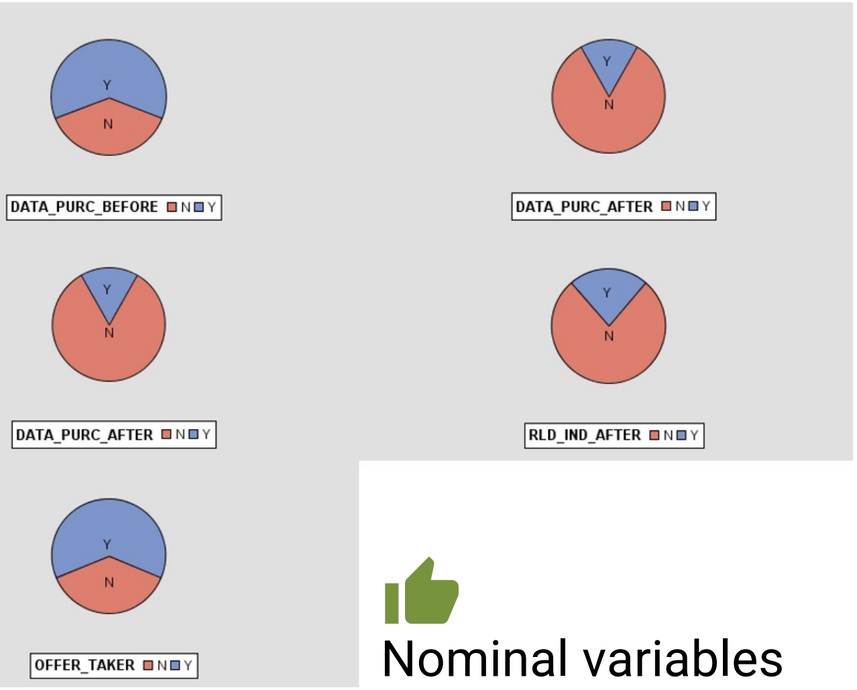
Univariate Analysis | Boxplot



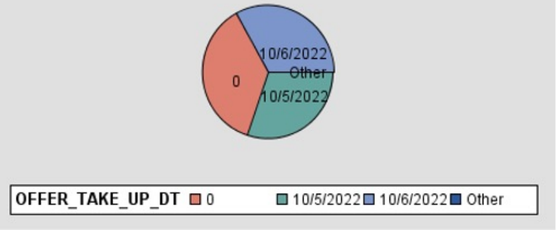
- No missing values
- All right-skewed distribution
- Outliers and noisy data detected

Explore

Univariate Analysis | Pie chart



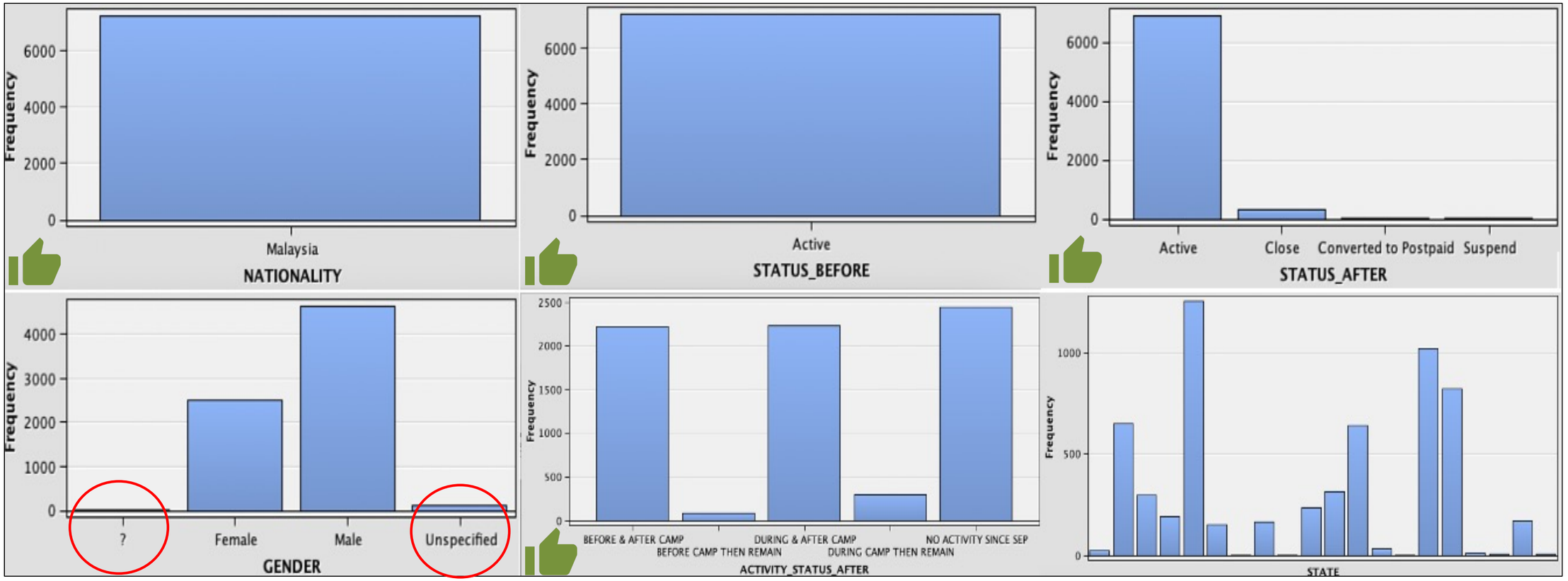
Nominal variables



Many missing values
Wrongly tagged as nominal

Explore

Univariate Analysis | Bar Chart

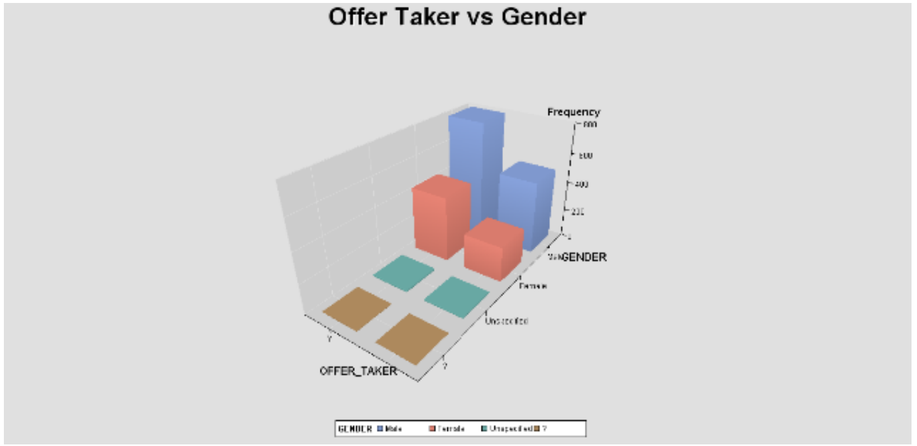


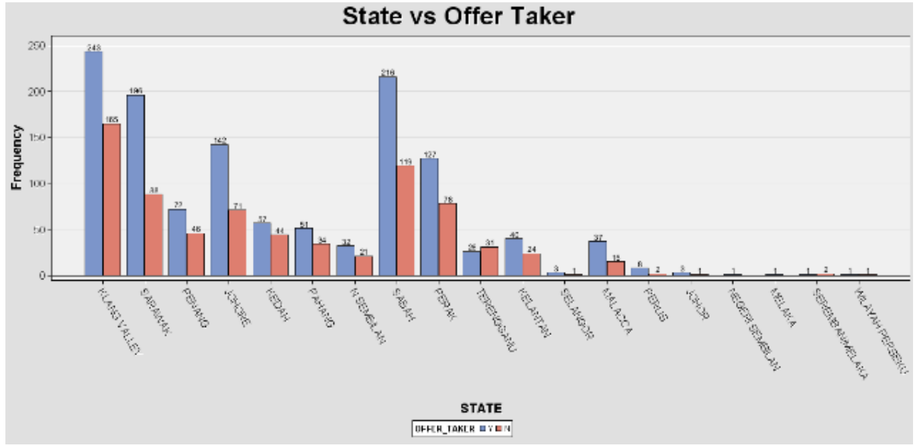
Inconsistent value:
'?' and 'unspecified' is same

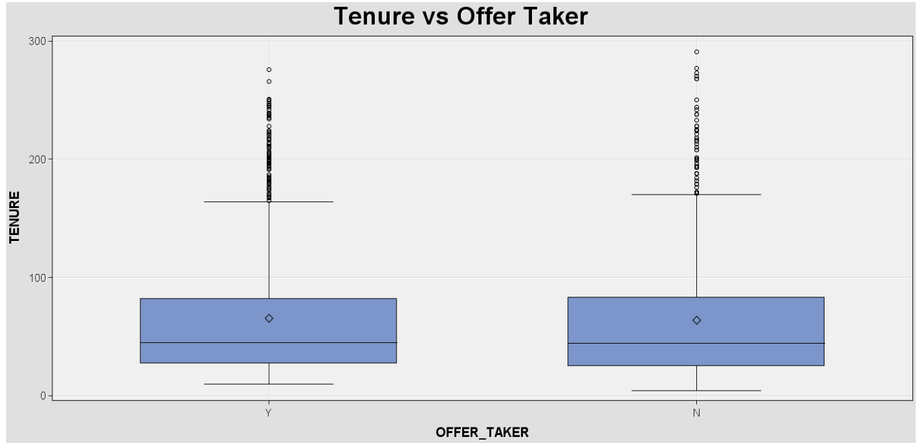
Inconsistent value:
more than 13 states

Explore

Bivariate Analysis



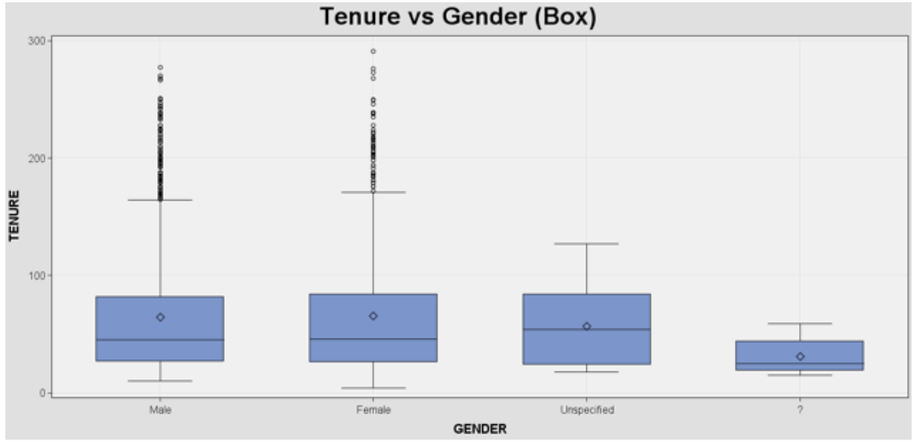


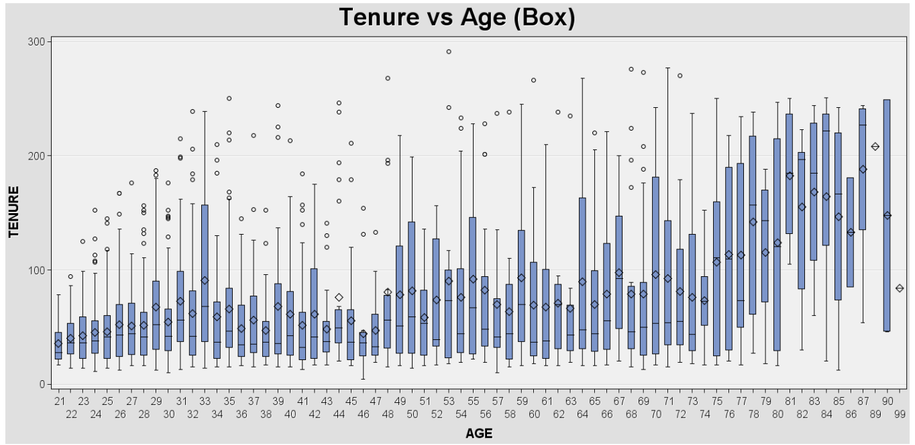


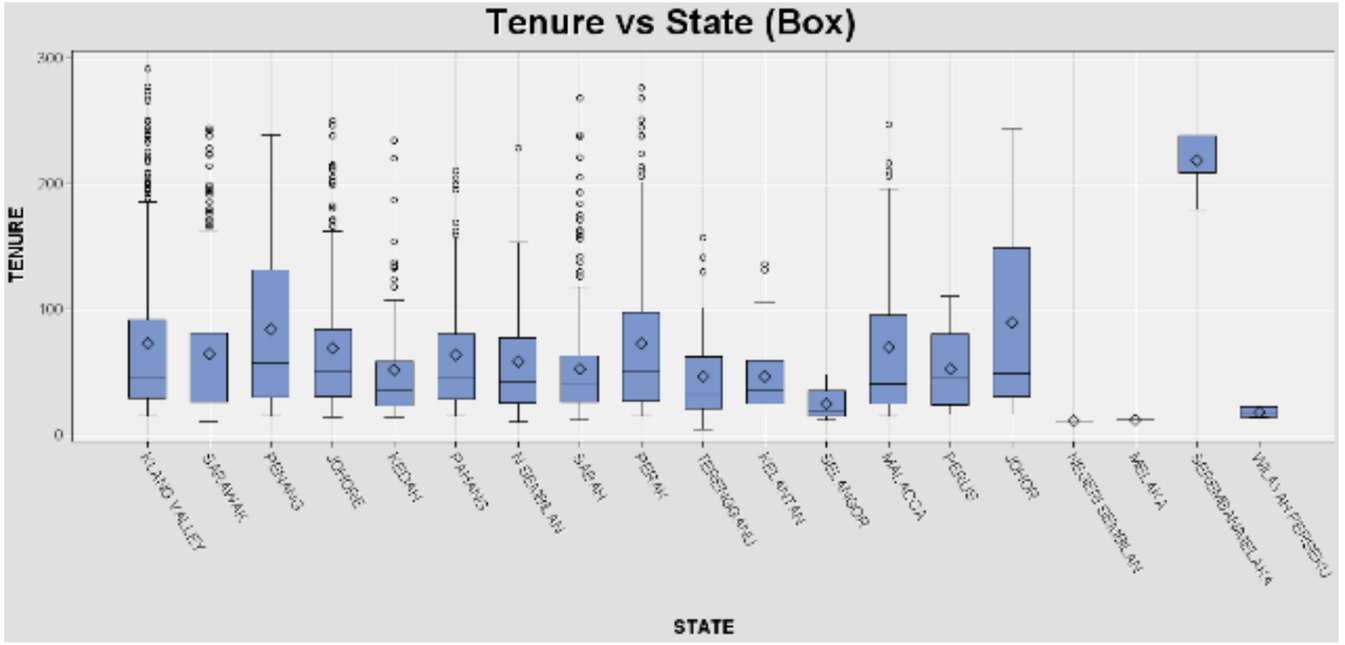
- Males likely to join campaign, most participants from Klang Valley, median age of 34 for both, median tenure of ~45

Explore

Bivariate Analysis



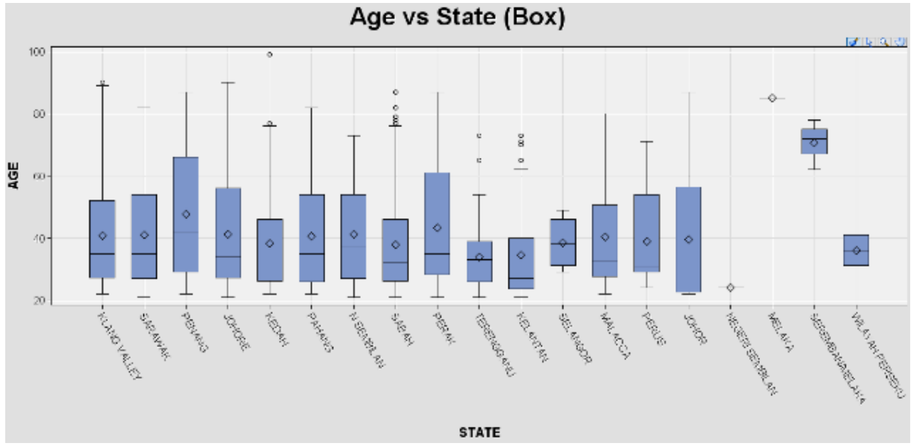


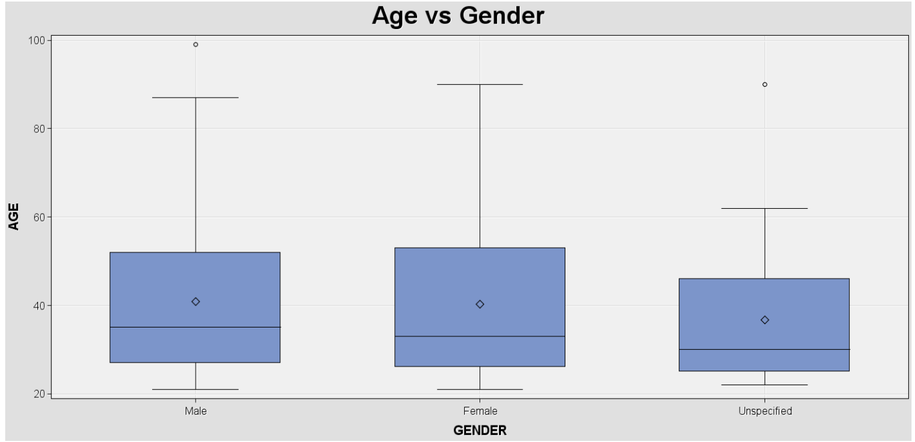


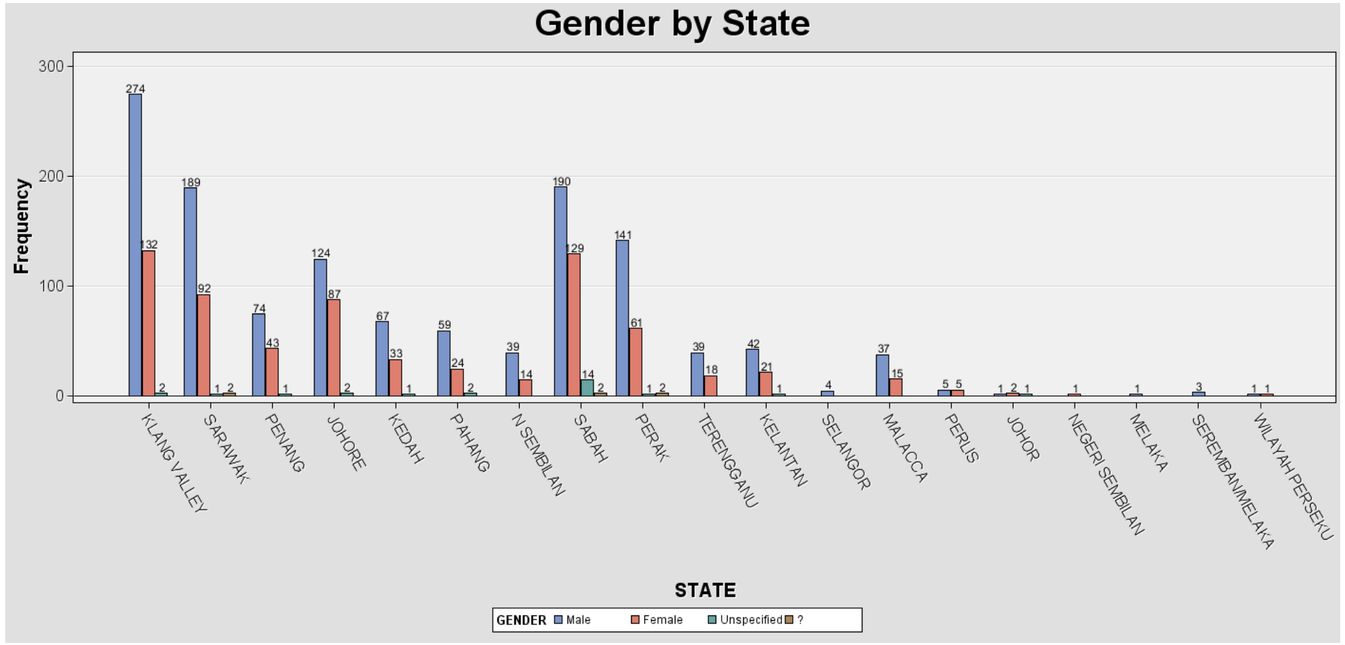
- Similar tenure distribution for both male & female
- Younger age = shorter tenure
- Varying tenure period in all states – surprisingly short tenures in Kedah

Explore

Bivariate Analysis



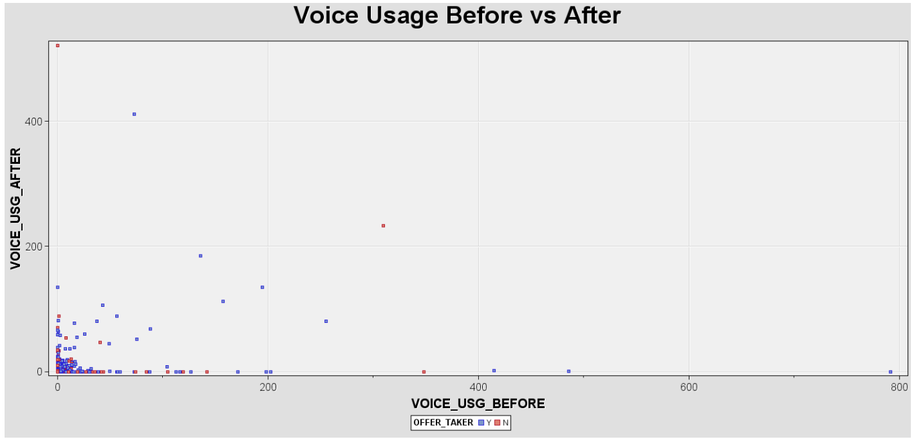


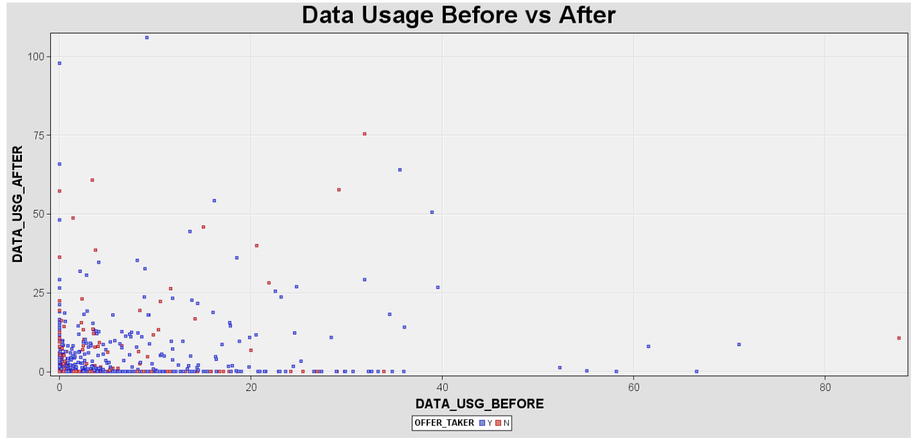


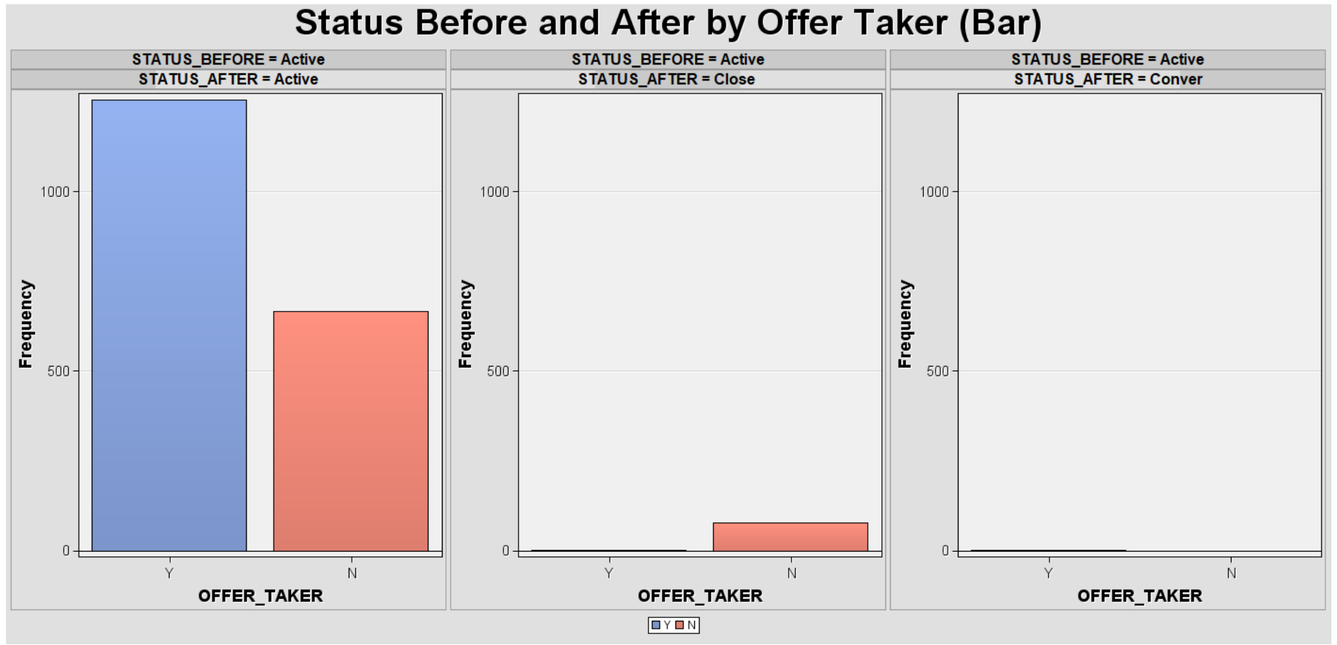
- Exceptionally young group of participants in Terengganu and Kelantan
- Similar median age for male and female (~34)
- Sabah has the most customers with undisclosed gender

Explore

Multivariate Analysis



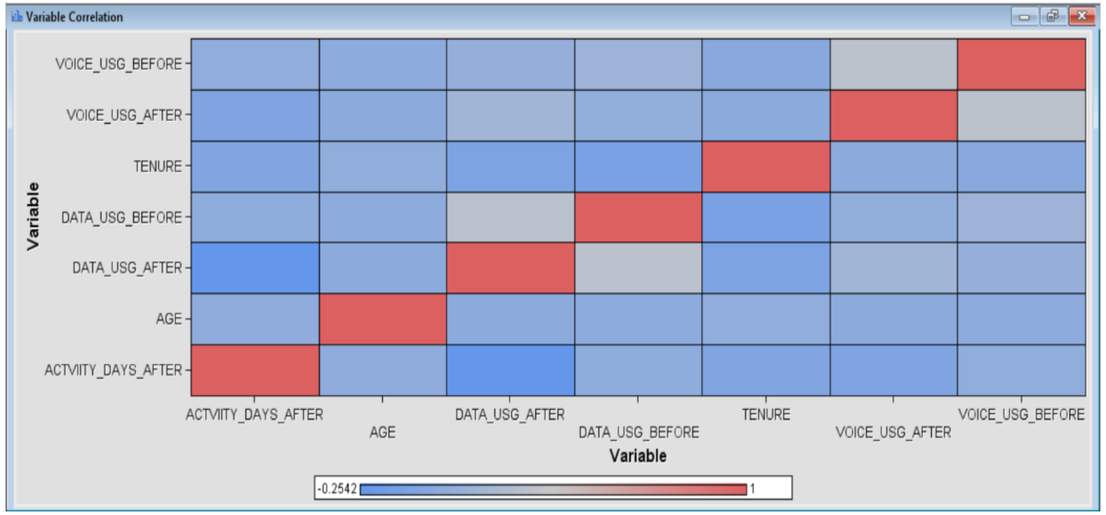


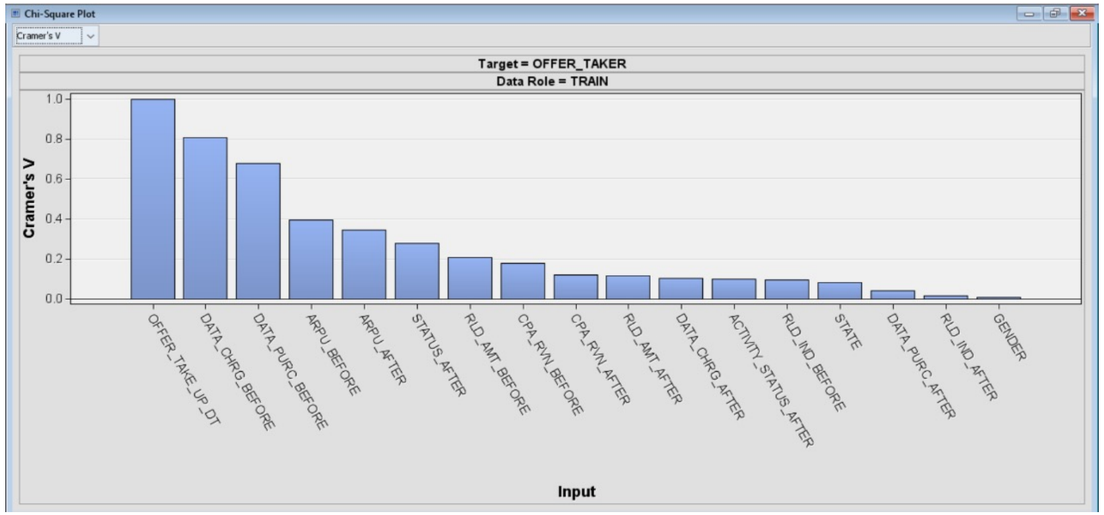


- The data and voice usage before and after campaign looks dispersed and exhibit no obvious relationship
- Takers more likely to remain active while non-takers more likely to churn

Explore

Correlation Analysis





For interval variables:

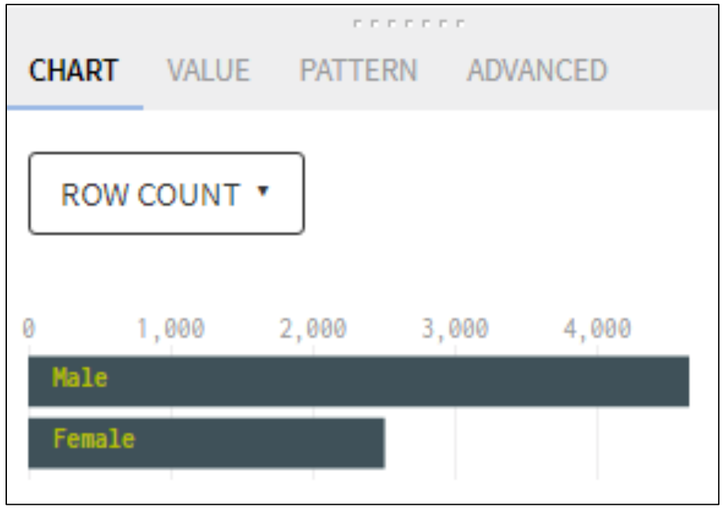
- No correlation value > 0.9
- No variables were removed

For nominal variables:

- OFFER_TAKE_UP_DT showed a prefect association

Modify

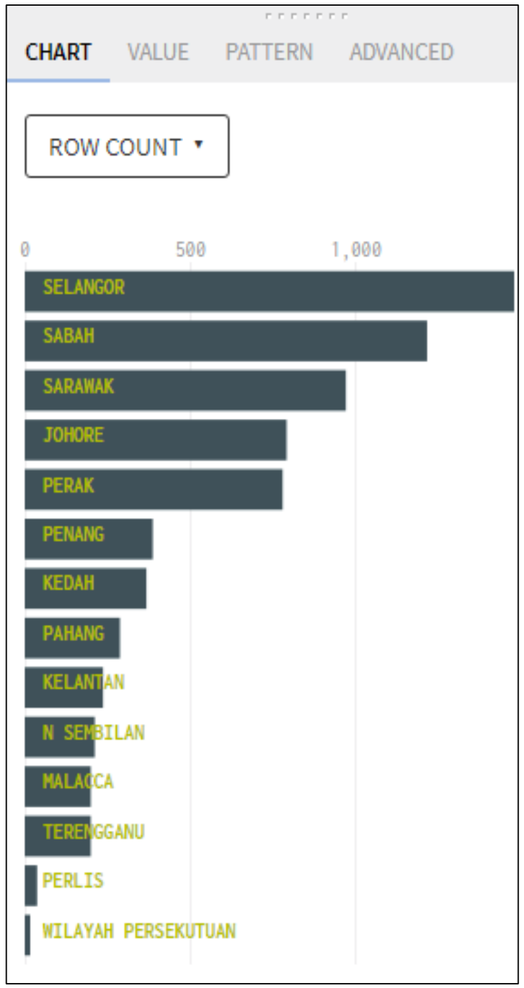
Modifying Inconsistent Data



GENDER:

- Initial: Male, Female, Unspecified, and ‘?’
- Current: Male, Female

* ‘Unspecified’ and ‘?’ were removed

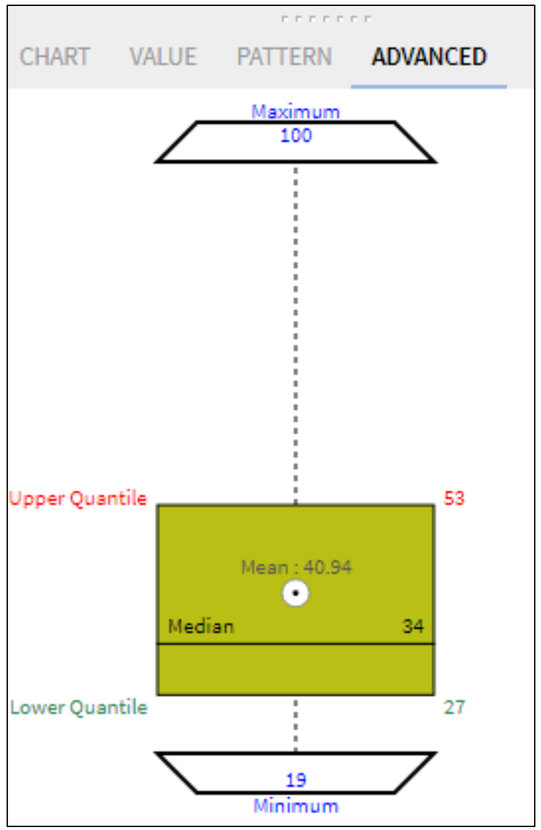


STATE:

- Initial: Naming consist of mixture of Malay and English
- Current: English naming

Modify

Modifying Intentional Data



AGE:

- Records with values of ‘-9999’ are removed
- Boxplot shows no outlier were detected

Modify

Modifying Incomplete Data

Variables involved:

- ARPU_BEFORE
- ARPU_AFTER
- CPA_RVN_BEFORE
- CPA_RVN_AFTER
- DATA_CHRG_BEFORE
- DATA_CHRG_AFTER
- RLD_AMT_BEFORE
- RLD_AMT_AFTER
- OFFER_TAKE_UP_DT

- Value ‘?’ is replaced with 0
- Data type converted from string to integer
- OFFER_TAKE_UP_DT remained as string as non-offer taker will not have date

Modify

Modifying Noisy Data (Outlier)

Variables involved:

- ARPU_BEFORE
- ARPU_AFTER
- CPA_RVN_BEFORE
- CPA_RVN_AFTER
- DATA_CHRG_BEFORE
- DATA_CHRG_AFTER
- RLD_AMT_BEFORE
- RLD_AMT_AFTER
- DATA_USG_BEFORE
- DATA_USG_AFTER
- VOICE_USG_BEFORE
- VOICE_USG_AFTER
- AGE
- TENURE

Outliers were not removed as the extreme values are legitimate observations that were part of the sample

Modify

Examining Exported Data for Modelling

Interval Variable Summary Statistics

Variable	Label	Missing	N	Minimum	Maximum	Mean	Standard Deviation	Skewness	Kurtosis
AGE	AGE	0	7157	19	100.00	40.9409	17.2254	0.8707	-0.43
ARPU_BEFORE	ARPU_BEFORE	0	7157	0	1505.28	10.4369	68.9795	14.7054	249.71
CPA_RVN_BEFORE	CPA_RVN_BEFORE	0	7157	0	1500.00	6.5504	74.0595	14.4383	229.53
DATA_CHRG_BEFORE	DATA_CHRG_BEFORE	0	7157	0	105.00	9.6490	9.8411	1.1641	3.30
DATA_USG_BEFORE	DATA_USG_BEFORE	0	7157	0	218.14	2.3890	7.6635	9.7988	181.16
RLD_AMT_BEFORE	RLD_AMT_BEFORE	0	7157	0	1955.00	12.6985	75.7843	15.2129	273.03
TENURE	TENURE	0	7157	3	296.00	63.9240	53.6474	1.6523	2.28
VOICE_USG_BEFORE	VOICE_USG_BEFORE	0	7157	0	2451.75	5.0188	53.1192	28.1807	1015.13

Class Variable Summary Statistics

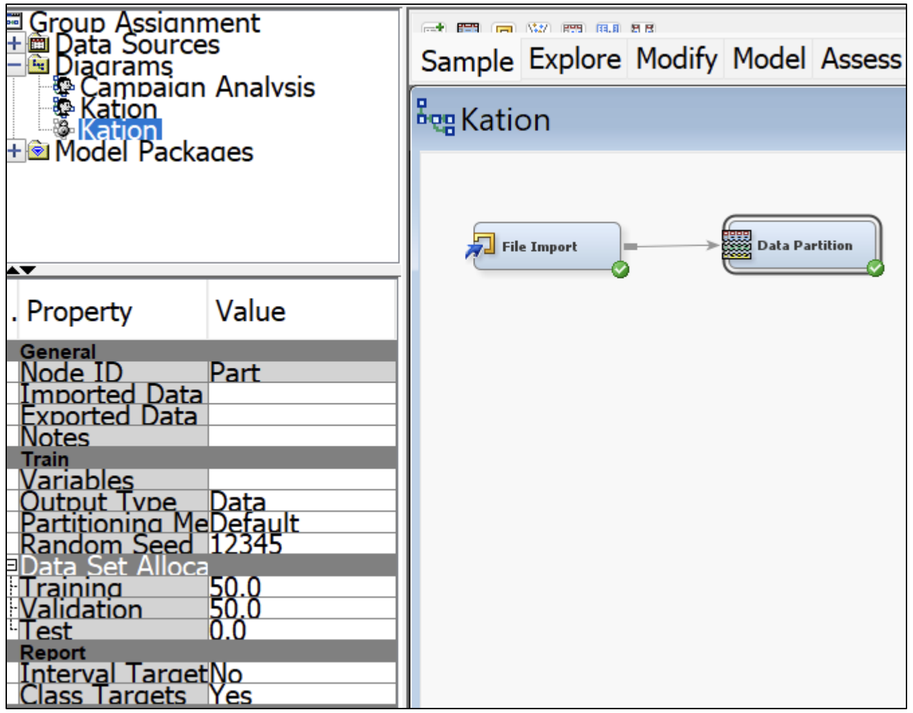
Variable	Label	Type	Number of Levels	Missing
DATA_PURC_BEFORE	DATA_PURC_BEFORE	C	2	0
GENDER	GENDER	C	2	0
OFFER_TAKER	OFFER_TAKER	C	2	0
RLD_IND_BEFORE	RLD_IND_BEFORE	C	2	0
STATE	STATE	C	14	0

After modification:

- No missing values were found
- Number of levels for class variables were correct

Modify

Data Partition



Dataset was split into 50% to training set and 50% to validation set

Summary Statistics for Class Targets

Data=DATA

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
OFFER_TAKER	.	N	2658	37.1385	OFFER_TAKER
OFFER_TAKER	.	Y	4499	62.8615	OFFER_TAKER

Data=TRAIN

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
OFFER_TAKER	.	N	1329	37.1333	OFFER_TAKER
OFFER_TAKER	.	Y	2250	62.8667	OFFER_TAKER

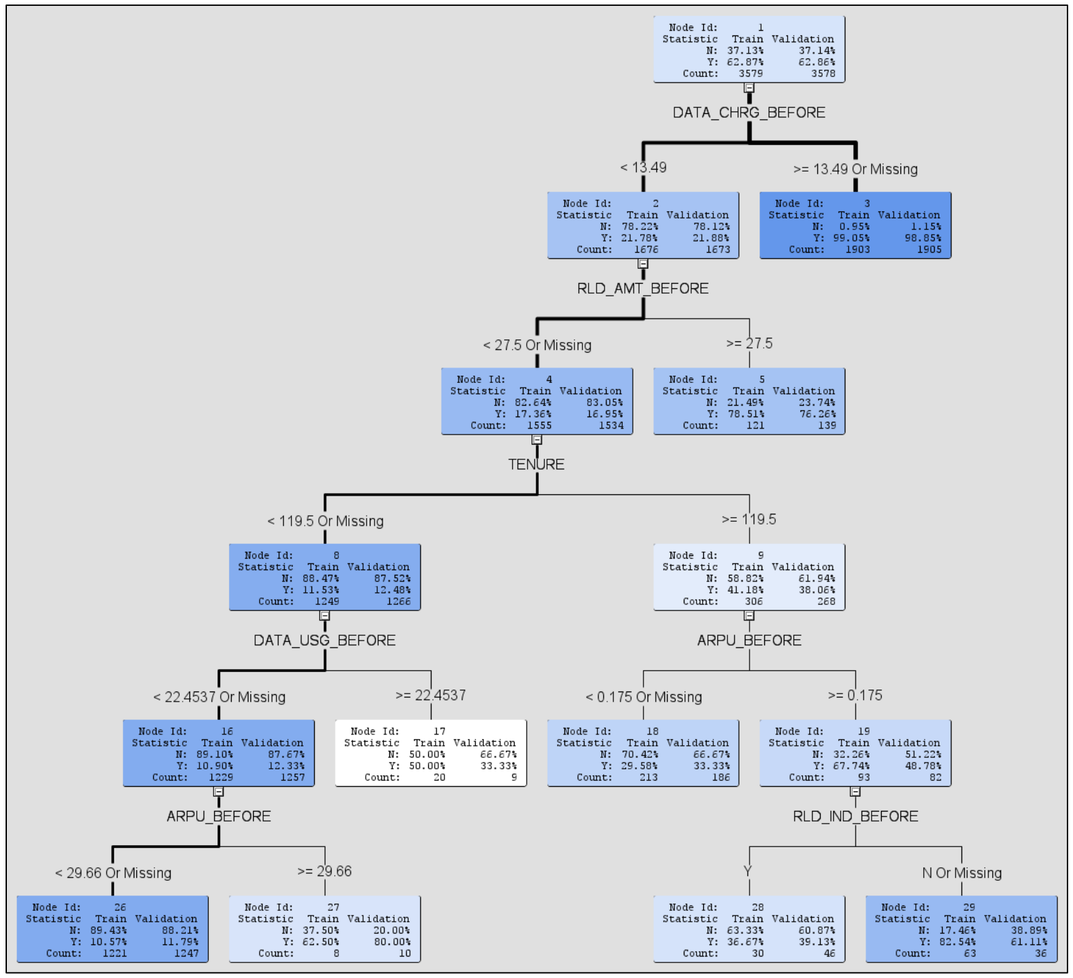
Data=VALIDATE

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
OFFER_TAKER	.	N	1329	37.1437	OFFER_TAKER
OFFER_TAKER	.	Y	2249	62.8563	OFFER_TAKER

Summary statistics of OFFER_TAKER

Model

Decision Tree



Campaign offer takers observation:

- Tends to purchase data plan with charge ≥ RM 13.49 before campaign
- Tends to reload ≥ RM 27.50 although they do not purchase extra data plan before campaign
- Have a longer tenure and higher ARPU

Model

Decision Rules & Feature Importance

- if DATA_CHRG_BEFORE >= 13.49 or MISSING then OFFER_TAKER=Y
- if RLD_AMT_BEFORE >= 27.5 AND DATA_CHRG_BEFORE < 13.49 then OFFER_TAKER=Y
- if TENURE < 119.5 or MISSING AND RLD_AMT_BEFORE < 27.5 or MISSING AND DATA_USG_BEFORE >= 22.4537 AND DATA_CHRG_BEFORE < 13.49 then OFFER_TAKER=Y
- if TENURE >= 119.5 AND RLD_AMT_BEFORE < 27.5 or MISSING AND DATA_CHRG_BEFORE < 13.49 AND ARPU_BEFORE < 0.175 or MISSING then OFFER_TAKER=N
- if TENURE < 119.5 or MISSING AND RLD_AMT_BEFORE < 27.5 or MISSING AND DATA_USG_BEFORE < 22.4537 or MISSING AND DATA_CHRG_BEFORE < 13.49 AND ARPU_BEFORE < 29.66 or MISSING then OFFER_TAKER=N
- if TENURE < 119.5 or MISSING AND RLD_AMT_BEFORE < 27.5 or MISSING AND DATA_USG_BEFORE < 22.4537 or MISSING AND DATA_CHRG_BEFORE < 13.49 AND ARPU_BEFORE >= 29.66 then OFFER_TAKER=Y
- if TENURE >= 119.5 AND RLD_IND_BEFORE IS ONE OF: Y AND RLD_AMT_BEFORE < 27.5 or MISSING AND DATA_CHRG_BEFORE < 13.49 AND ARPU_BEFORE >= 0.175 then OFFER_TAKER=N
- if TENURE >= 119.5 AND RLD_IND_BEFORE IS ONE OF: N or MISSING AND RLD_AMT_BEFORE < 27.5 or MISSING AND DATA_CHRG_BEFORE < 13.49 AND ARPU_BEFORE >= 0.175 then OFFER_TAKER=Y

Rules of Decision Tree model

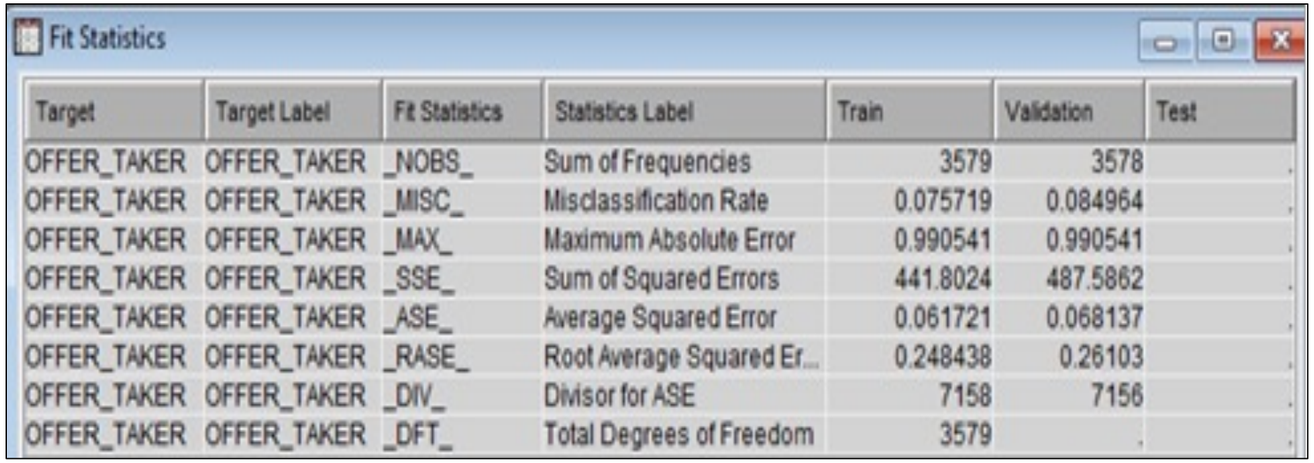
Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
DATA_CHRG_BEFORE	DATA_CHRG...	1	1.0000	1.0000	1.0000
RLD_AMT_BEFORE	RLD_AMT_B...	1	0.2809	0.2912	1.0367
TENURE	TENURE	1	0.2015	0.1636	0.8120
ARPU_BEFORE	ARPU_BEFO...	2	0.1475	0.0911	0.6178
RLD_IND_BEFORE	RLD_IND_BE...	1	0.0896	0.0652	0.7273
DATA_USG_BEFORE	DATA_USG_...	1	0.0752	0.0021	0.0285
AGE	AGE	0	0.0000	0.0000	.
DATA_PURC_BEFORE	DATA_PURC...	0	0.0000	0.0000	.
VOICE_USG_BEFORE	VOICE_USG_...	0	0.0000	0.0000	.
CPA_RVN_BEFORE	CPA_RVN_B...	0	0.0000	0.0000	.
STATE	STATE	0	0.0000	0.0000	.
GENDER	GENDER	0	0.0000	0.0000	.

- DATA_CHRG_BEFORE, RLD_AMT_BEFORE, TENURE, ARPU_BEFORE, RLD_IND_BEFORE and DATA_USG_BEFORE are the most important variables
- Can be utilized on next campaign to predict potential customers to take up the offer

Assess

Fit Statistics | Confusion Matrix



Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
OFFER_TAKER	OFFER_TAKER	_NOBS_	Sum of Frequencies	3579	3578	.
OFFER_TAKER	OFFER_TAKER	_MISC_	Misclassification Rate	0.075719	0.084964	.
OFFER_TAKER	OFFER_TAKER	_MAX_	Maximum Absolute Error	0.990541	0.990541	.
OFFER_TAKER	OFFER_TAKER	_SSE_	Sum of Squared Errors	441.8024	487.5862	.
OFFER_TAKER	OFFER_TAKER	_ASE_	Average Squared Error	0.061721	0.068137	.
OFFER_TAKER	OFFER_TAKER	_RASE_	Root Average Squared Er...	0.248438	0.26103	.
OFFER_TAKER	OFFER_TAKER	_DIV_	Divisor for ASE	7158	7156	.
OFFER_TAKER	OFFER_TAKER	_DFT_	Total Degrees of Freedom	3579	.	.

Fit Statistics:

- No overfitting
- Very low misclassification rate
  - 7.6% for training set and 8.5% for validation set
- Average squared error is close to 0

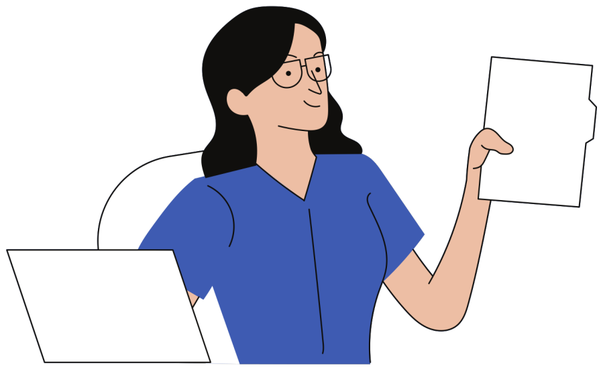
Metrics	Accuracy	Precision	Recall	Specificity	F1 measure
(Train)	92.4%	96.8%	91.0%	94.9%	94.0%
(Validate)	91.5%	96.3%	90.7%	94.2%	93.0%

Confusion Matrix:

- Decision Tree has a great performance in differentiating the campaign offer takers and non-offer takers.

Conclusion

Evaluating Telco Campaign Performance and Predicting Campaign Offer Takers are conducted using Data Mining Techniques and SEMMA method.



01

To assess the effectiveness of “Right Planning” pilot campaign

- 4573 takers (63% opt in rate)
- 2921 takers remain active after campaign (64% active rate)
- Moderately success

02

To identify campaign takers’ profile

- Target group is sample of Kation’s prepaid Malaysian
- Higher opt in rate among male
- Age group between 22 - 36 years old with tenure > 1 year
- Mostly from Klang Valley, then Sabah and Sarawak

03

To predict campaign takers based on usage and revenue behavior

- Offer takers tends to:
  - Purchase data plan with charge ≥ RM 13.49
  - Reload ≥ RM 27.50
  - Have a long tenure and high ARPU
- Decision Tree performed extremely good in classifying the campaign offer takers with non-offer takers.

That's a wrap!

Thank you.

12th Jan 2023