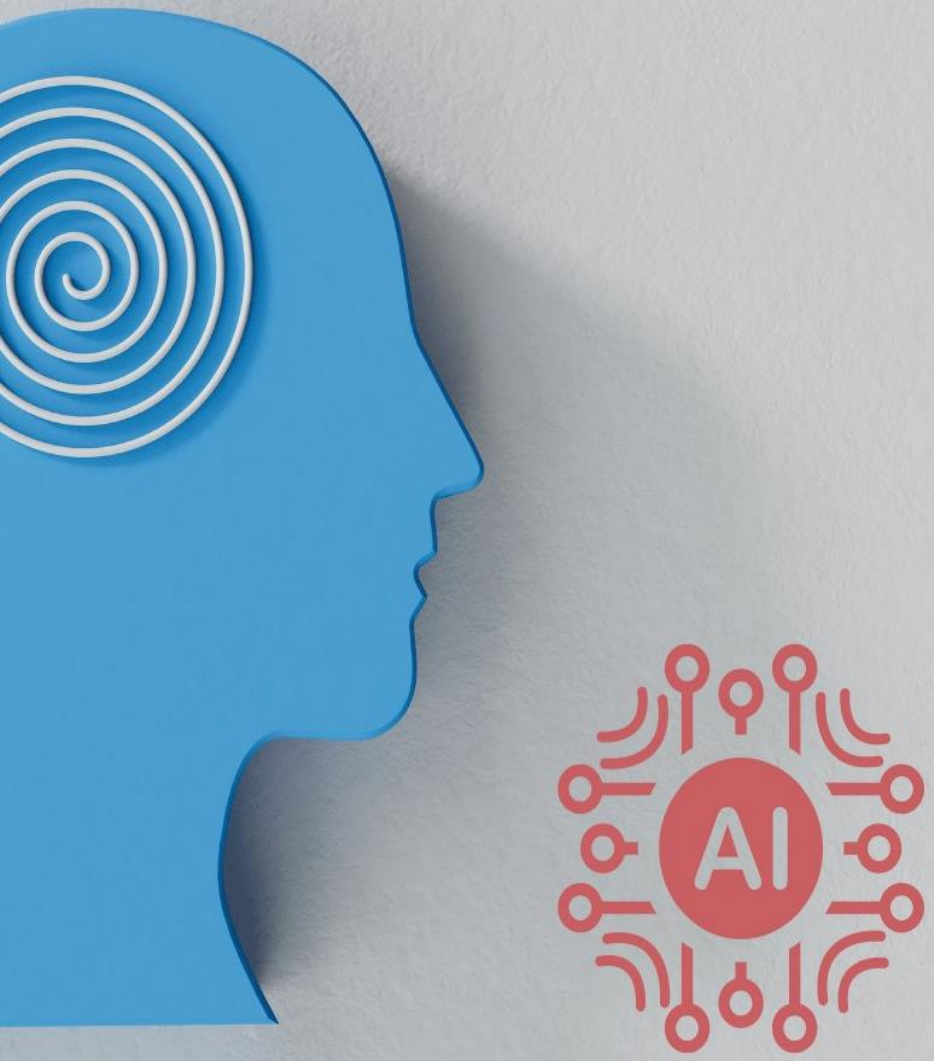


X

Telco Customer Churn Tendency

Revou x Kampus Merdeka Data Analytics Mentor Application

by Rayhan Narawangsa

January 2024

Background and Objective

Background and Objective

- Background: Our telecom company have annual churn rate of 10-25%. This become a huge problem since acquiring new customers is 5-10 more pricey that retaining an existing one.
- Objective: **Determine root cause** that make customer churn, and **provide recommendations** for customer portofolio retention.

Scope of Work

Driver
Project Initiatives by Strategy Team (Done)

Approver
Resource Approval by Head of Consumer Business (Done)

Responsible
Root Cause and Action Plan Determination by Analytics Team (On Progress)

Consulted
Customer Pitching and Strategic Execution by Data Analytics and Campaign Team

Informed
Final Approval and Legal Checking by Campaign and Business Team

Analytics Approach

1 **Determine Objective and Scope of Work.**
Product Usage, Revenue – Cost Structure, Usage Behavior.

2 **Data Cleansing and Exploratory Enalysis.**
Defining target variable (Churn Tagging), Handling missing value, data engineering.

3 **Root Cause Determination.**
Univariate – Multivariate Analysis, Hyphotesis test, Generate issue tree and customer sizing

4 **Generating Recommendations and Action Plan.**
Making structured action plan to retain churning customers, Coordinate with related stekholders.

Dataset Understanding

	Acronyms	Descriptions
1	MOBILE_NUMBER	Customer phone number
	CIRCLE_ID	Telecom circle area to which the customer belongs to
	LOC	Local calls - within same telecom circle
	STD	STD calls - outside the calling circle
	IC	Incoming calls
2	OG	Outgoing calls
	T2T	Operator T to T, i.e. within same operator (mobile to mobile)
	T2M	Operator T to other operator mobile
	T2O	Operator T to other operator fixed line
	T2F	Operator T to fixed lines of T
	T2C	Operator T to it's own call center
3	ARPU	Average revenue per user
	MOU	Minutes of usage - voice calls
	AON	Age on network
	PCK	Prepaid service schemes called - PACKS
	NIGHT	Scheme to use during specific night hours only
	MONTHLY	Service schemes with validity equivalent to a month
	SACHET	Service schemes with validity smaller than a month
	FB_USER	Service scheme to avail services of Facebook or similar sites
3	VBC	Volume based cost
	ONNET	All kind of calls within the same operator network
	OFFNET	All kind of calls outside the operator T network
	ROAM	Indicates that customer is in roaming zone during the call

Acronyms	Descriptions
SPL	Special calls
ISD	ISD calls
RECH	Recharge
NUM	Number
AMT	Amount in local currency
MAX	Maximum
2 DATA	Mobile internet
3G	3G network
AV	Average
VOL	Mobile internet usage volume (in MB)
2G	2G network
*.6	KPI for the month of June
*.7	KPI for the month of July
*.8	KPI for the month of August
*.9	KPI for the month of September

1. Analysis will be on mobile_number level as primary key. Each are complimented by demography data in Age of Network and Circle_ID (Area Code).

2. Types of product offered from T Network are General Calls and Internet packs. The data includes usage detail sush as usage volume, channel, and packs offered.

3. The data provided includes financial aspect as in Average Revenue and Volume Based Cost per mobile_number.

Data Cleansing – Missing Value Findings

- Massive portion of rows with missing value are found in several key features.
- Most features with highest percentage is related to recharge data, revenue, and packs.
- General approaches of handling missing value are applied to anticipate this case.

Columns	Missing Value (%)	Columns	Missing Value (%)
date_of_last_rech_data	75%	av_rech_amt_data	74%
total_rech_data	75%	arpu	74%
max_rech_data	75%	night_pck_user	74%
count_rech	75%	fb_user	74%

Cases of Handling Missing Value

In facing massive amount of data, we need **efficient approach** to handle missing values. Used Approach: Data Reduction and Replacement using Statistic Imputation.

Data Reduction

- Drop columns: Used for targeted columns with huge portion of missing value (used threshold > 40%), and are less relevant to the analysis.
- Targeted Columns: Last_date_of_month.

Statistics Imputation

- Replace with number 0**:** Targeted column have clear rules with high confidence level in changing data distribution (eg. fb_user).
- Replace with Median: to effectively replace missing values in huge number of columns (eg std_ic_moud, etc).

Illustration

Mobile Number	Last_date_of_month	Circle_id	fb_user_6	fb_user_7	total_ic_6	total_ic_7	std_ic_mou
708+++++	31/06	1248	1	1	87.1	34.9	NaN
701+++++	NaN	1273	0	NaN	280.08	216.61	4.94
702+++++	31/06	NaN	0	0	532.8	NaN	458.5
Action	**Drop Column**		**Replace with 0**		**Replace with Median**		

Exploratory Analysis – Defining Churn Variables and Generate Additional Feature

Churn Variables

Customers (mobile_number) who have shown no activity, including both incoming and outgoing calls, as well as internet usage, during the month of September (month 9). Churn Value: **1 for customer churn, 0 for customer non churn.**

Illustration

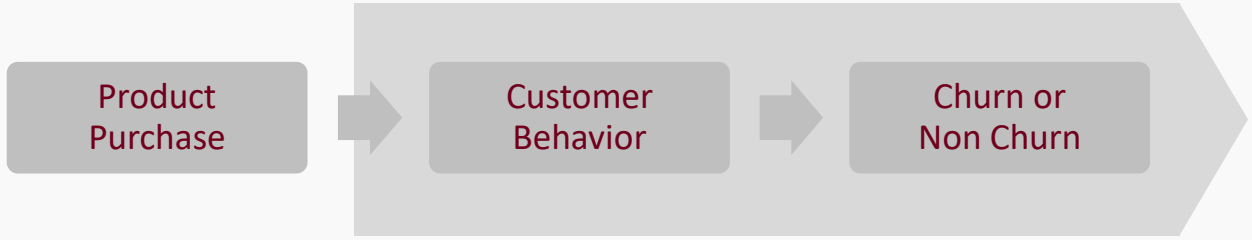
	Target Period				Target Period				
Mobile Number	Inet Usage Month 6	Inet Usage Month 7	Inet Usage Month 8	Inet Usage Month 9	Calls Usage Month 6	Calls Usage Month 7	Calls Usage Month 8	Calls Usage Month 9	Churn
708+++++	54	39	12	0	543	492	532	548	0
701+++++	32	25	41	38	679	432	125	0	0
703+++++	21	17	6	0	320	179	53	0	1

Customer Sizing

Status	# Customer	% Customer
Churn	10,191	10.2
Non Churn	89,808	89.8
Total	**99,999**	**100**

Additional Feature

Customer Journey



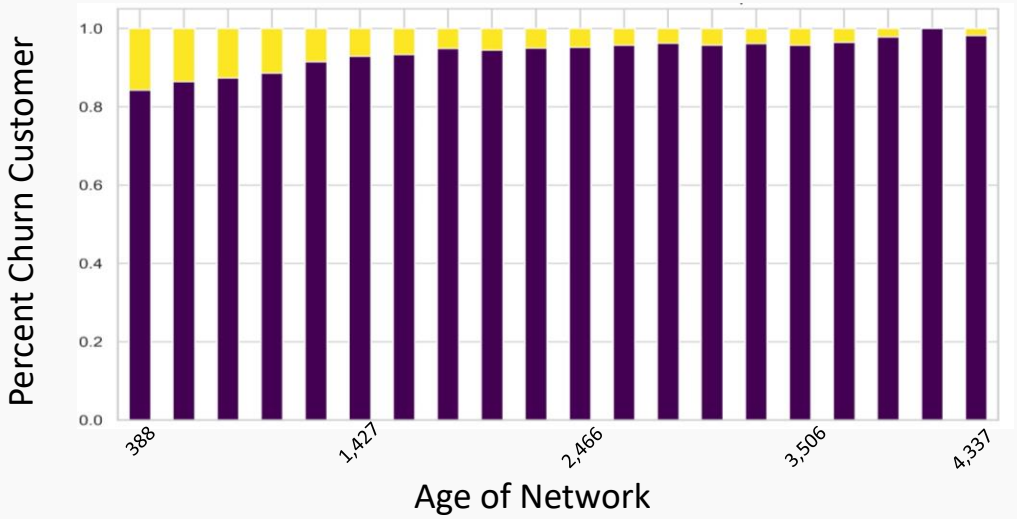
Additional Feature

What data or feature will be best to define customer behavior pattern and changes?

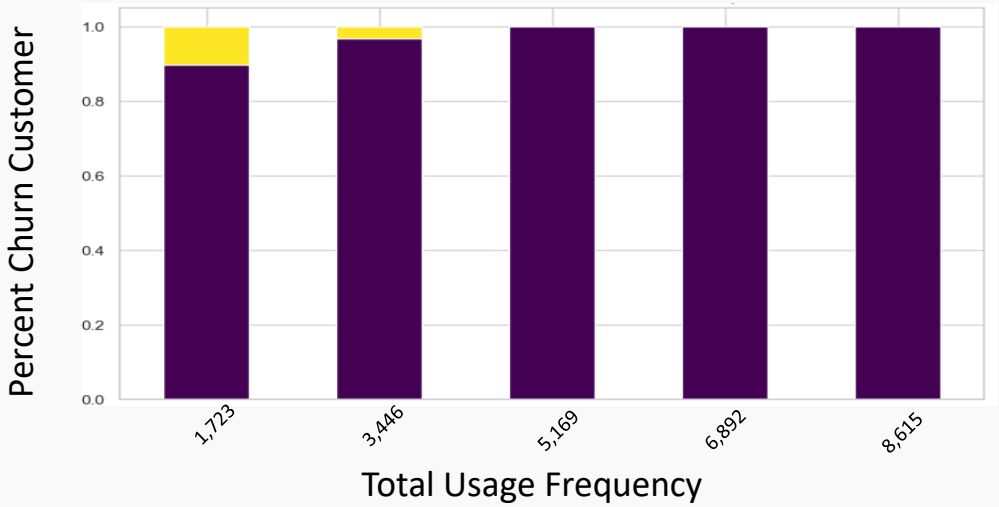
- Behaviour Sum first 3 month
- Behaviour Trend first 3 month
- Behaviour Average first 3 month
- Years Age of Network

Exploratory Analysis and Hipothesis Determination

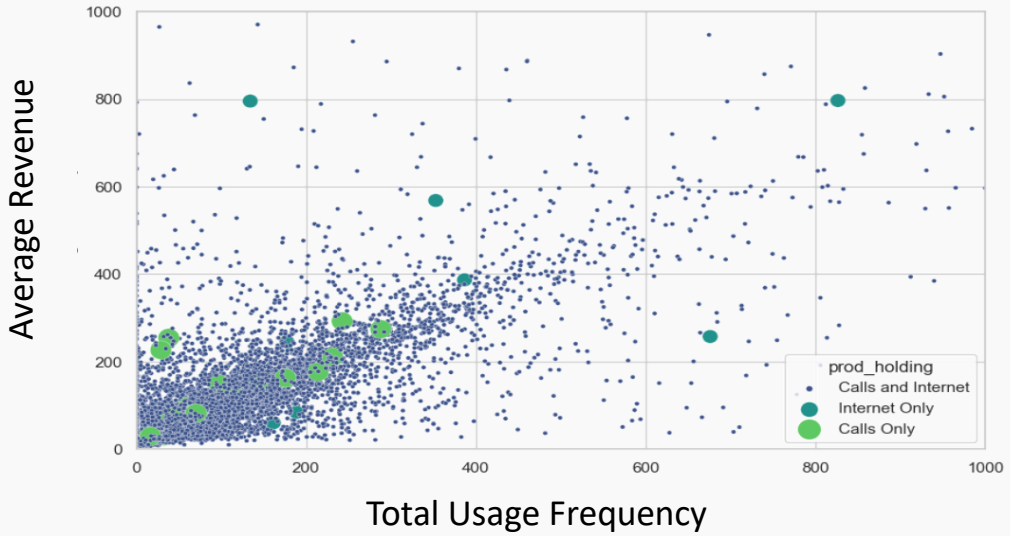
Churn Sizing vs Age of Network (AON) 1



Churn Sizing vs Usage Frequency 2



Churn Sizing vs Monetary vs Products Usage 3



Non Churn Customers Churn Customers

* Average Revenue = Total Revenue Calls and Internet
** Total Usage Frequency = Total Usage Calls and Internet After Normalized

Hyphotesis

1. Customers with low Age of Network are more likely to churn.
2. Customers with low product utilization are more likely to churn.
3. Retaining customers who hold product Internet only is more profitable than Calls only.

Churn Customer RFM Clustering

Variables (First 3 Month Behavior)		Cluster A (Champion)	Cluster B (Dynamic Achievers)	Cluster C (Balanced Explorers)	Cluster D (Efficient Subscriber)	Cluster E (Fresh Minimalist)
Customer Sizing		20,000 cust.	20,000 cust.	19,999 cust.	20,000 cust.	20,000 cust.
Churn Customer Sizing (1)		525 cust. (2.6%)	1,093 cust. (5.5%)	1,566 cust. (7.8%)	2,122 cust. (10.6%)	4,885 cust. (24.4%)
Age of Network		58 months	41 months	40 months	42 months	23 months
Recency	Gap from last Call Recharge	4 days	5 days	6 days	9 days	16 days
	Gap from last Inet Recharge (2)	12 days	34 days	49 days	57 days	60 days
Frequency*	Outgoing Calls Minutes (3)	500 minutes	403 minutes	313 minutes	193 minutes	121 minutes
	Incoming Calls Minutes	324 minutes	230 mintues	192 minutes	157 minutes	97 minutes
	2G Internet Volume Usage	148 mb	77 mb	25 mb	4 mb	0.5 mb
	3G Internet Volume Usage	461 mb	147 mb	29 mb	4 mb	0.8 mb
Monetary*	Revenue per Person from Calls	508 cu	349 cu	256 cu	188 cu	117 cu
	Revenue per Person from Inet	83 cu	23 cu	5 cu	5 cu	0.2 cu
Revenue x Churn Customer** (4)		310,275 cu	406,496 cu	408,726 cu	409,546 cu	572,522 cu

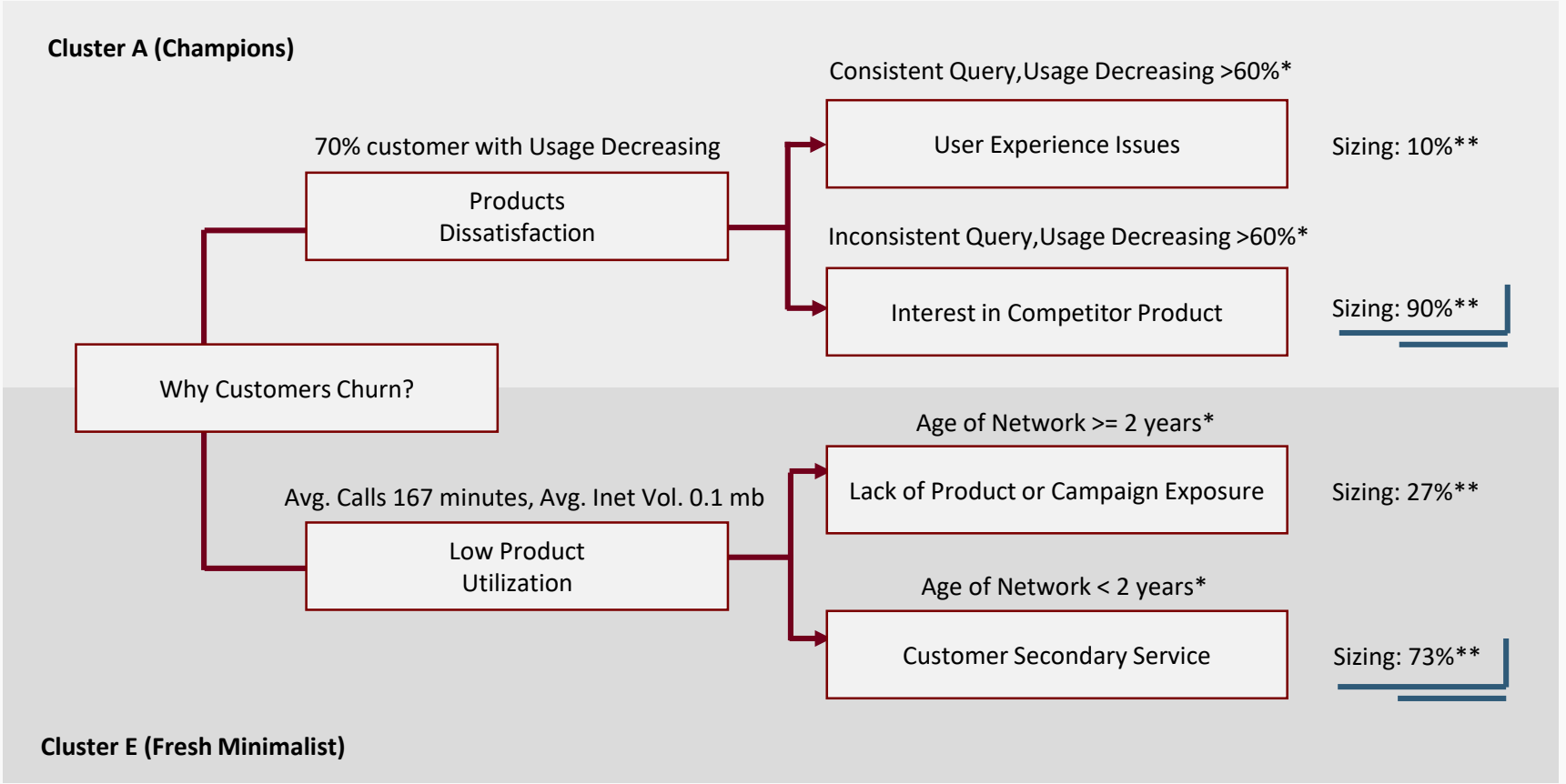
* Average first 3 month ** Total Revenue (Calls and Internet) times by Number of Churn Customer

Key Highlight

1. Based on churn percentage, Fresh Minimalist customers are more likely to churn.
2. Churn customers typically exhibit a more longer gap from their last recharge cycle to the current month (September).
3. Customer with low product usage is more likely to churn. Better product usage to determine churn tendency is internet.
4. Highest revenue lost are from cluster Fresh Minimalist.

Customer Churn Trigger – Issue Tree

Cluster A and E is chosen to determine root cause for its representative natures.



Definition

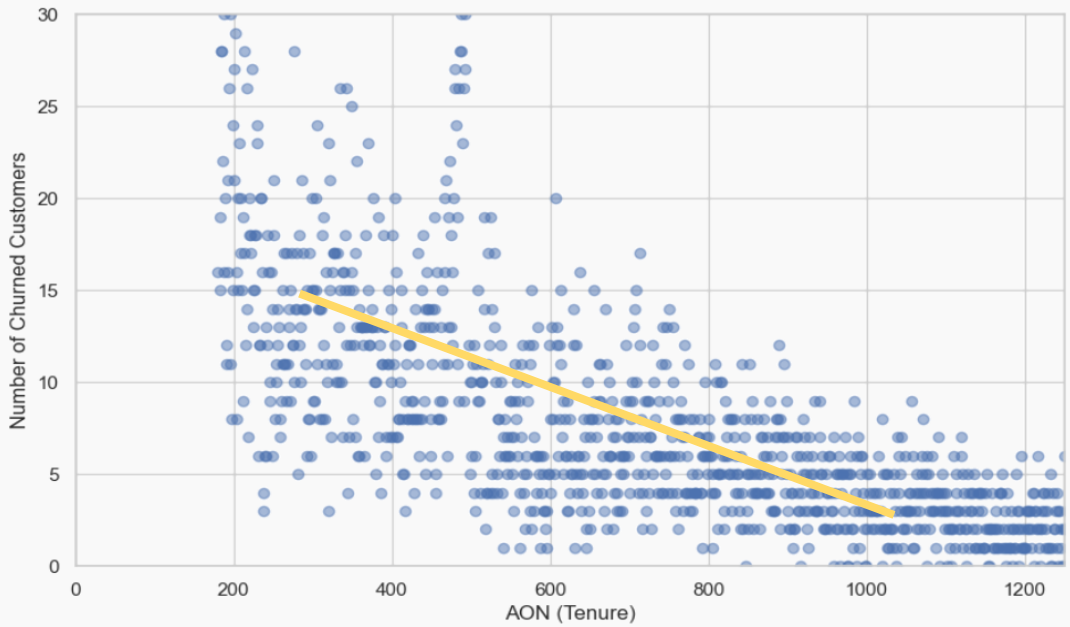
* Used threshold for customer sizing. Population is from Churning customer in each respective Cluster.
** Used to determine the main cause of churning between two primary types of Customer cluster (Champions and Fresh Minimalist).

Key Highlights

The two main cause of customer churning are Interest in Competitor Product and Non Focus Awareness from Customer

Strategy and Action Plan – Hyphothesis Test

Example Correlation Plot AON vs Churn Customers



Approach

- Customer distribution plot provide support on finding relationship between features
- In AoN vc Churn Customer case, plot suggest there is high correlation between age of network and number of churned customers.
- Logistic Regression is used to find the statistics prove to support hyphothesis.
- P-value shows from the relationship between Age of Network and Number of Churn Customer is <0.05, which surpass limit of 95% confidence level.
- Currelation plot and Logistic Regression is also generated for other hyphotheiss, which also shows similar result, suggesting that all 3 hyphothesis is correct:

Hyphothesis		Function Value	Logistic Regression Coefficient	P-Value	Status
Customers with low Age of Network are more likely to churn.		0.319784	-0.06263	0.011862	Correct
Customers with low product utilization are more likely to churn.		0.318215	-0.46827	0.023827	Correct
Retaining customers who hold product Internet only is more profitable than Calls only.	Internet Usage vs Churn Sizing	0.331627	-0.34162	0.024612	Correct
	Calls Usage vs Churn Sizing	0.332562	-0.34923	0.031592	

Strategy and Action Plan

1. Build promotion campaign to retain churning customers with higher product usage.

2. Build promotion campaign to transform new customers with low usage, into high usage customer.

3. Put more focus on churning customer with internet only product holding, with relatively higher monetary values.

Variables (First 3 Month Behavior)		Cluster A (Champion)	Cluster E (Fresh Minimalist)
Customer Sizing		20,000 cust.	20,000 cust.
Churn Customer Sizing		525 cust. (2.6%)	4,885 cust. (24.4%)
Age of Network		58 months	23 months
Recency	Gap from last Call Recharge	4 days	16 days
	Gap from last Inet Recharge	12 days	60 days 3
Frequency*	Outgoing Calls Minutes	500 minutes	121 minutes
	Incoming Calls Minutes	324 minutes	97 minutes
	2G Internet Volume Usage	148 mb	0.5 mb
	3G Internet Volume Usage	461 mb	0.8 mb
Monetary*	Revenue per Person from Calls	508 cu	117 cu
	Revenue per Person from Inet	83 cu 1	0.2 cu
Revenue x Churn Customer**		310,275 cu	572,522 cu 2

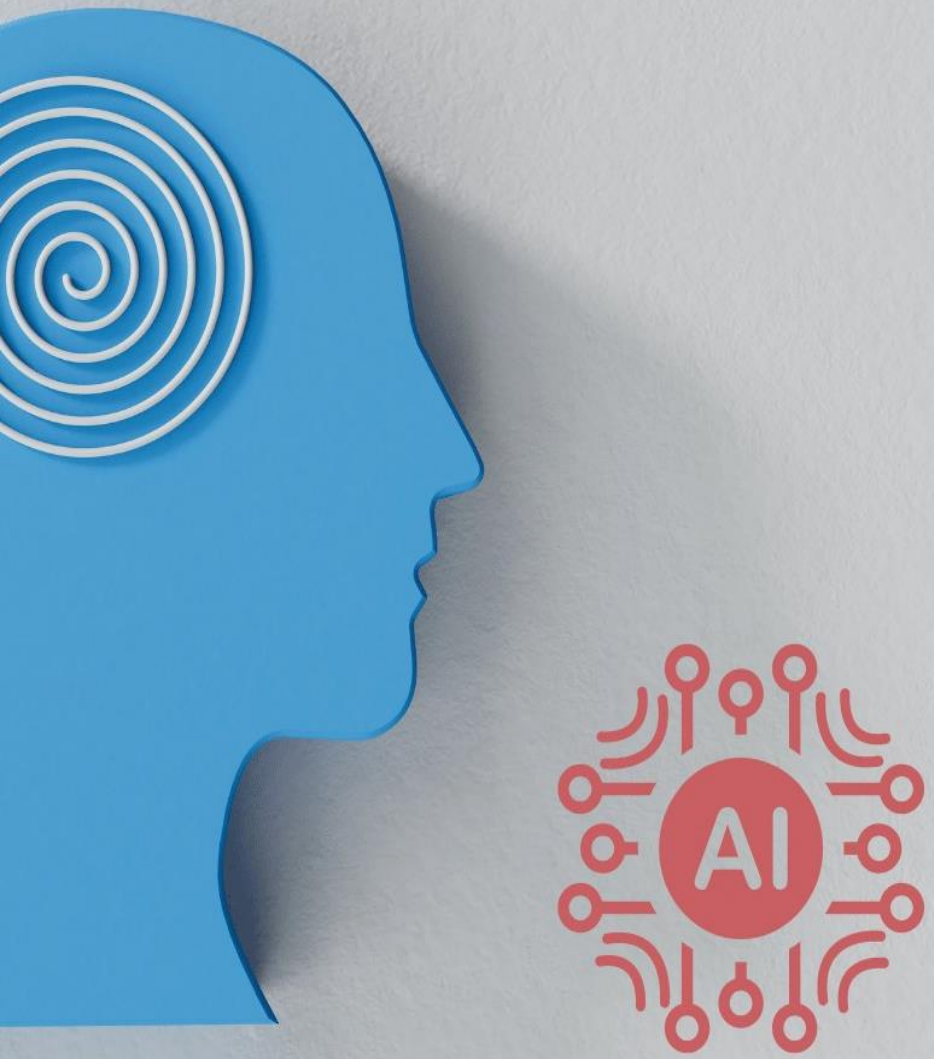
Next Action Plan

Root Cause and Action Plan Determination by Data Analytics Team (Done)

Customer Pitching and Strategic Execution by Data Analytics and Campaign Team (Next)

Final Approval and Legal Checking by Campaign and Business Team

Build machine learning model to predict Champions customer that will churn



X

Thank you

Revou x Kampus Merdeka Data Analytics Mentor Application

by Rayhan Narawangsa

January 2024