SCADS
Powered by StarCore

Applied Data Science for Customer Segmentation in CRM

23-24 November 2023

Supported by

StarCore analytics

IPB University
Bogor Indonesia

Department of Statistics

SCADS
Powered by StarCore

Excited?

This free course will be held in

2 Days

SCADS
Powered by StarCore

TWO DAYS Free Course

Day 1	Day 2
• Introduction to CRM • Introduction to Data Analytics and Customer Segmentation • Data Mart for Customer Segmentation • Hands-on Data Preparation and Data Exploration	• Introduction to Cluster Analysis • Concept of Cluster Analysis using K-means • Hands-on Customer Segmentation

SCADS
Powered by StarCore

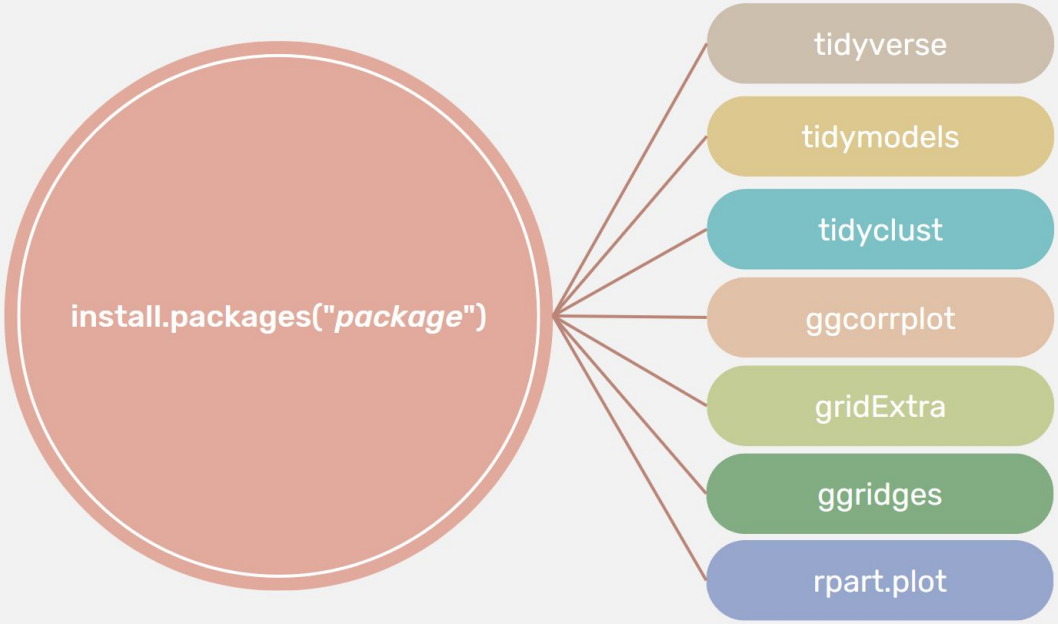
Aep Hidayatuloh

Head of Data Analytics

aep.hidayatuloh@starcore.co
https://starcore.co/scads

https://youtube.com/@aephidayatuloh
@aephidayatuloh
aephidayatuloh
aephidayatuloh

Software & Tools



Data & Code: https://github.com/aephidayatuloh/segmentation

SCADS
Powered by StarCore

Day 1

SCADS
Powered by StarCore

Table of contents

SCADS
Powered by StarCore

01

Customer Relationship Management

SCADS
Powered by StarCore

“All customers are not created equal”

Customer Relationship Management

CRM is a company-wide comprehensive strategy to build and nurture mutually benefit relationship with customers, through every touch-point, supported by technology, analytics, system and procedures in order to understand, to get close to, to gain trust of Customers by providing

the **Right Offer**

to the **Right Person**

at the **Right Time**

through the **Right Channel**

SCADS
Powered by StarCore

02

Data Analytics and Customer Segmentation

SCADS
Powered by StarCore

"Data analytics is the future, and the future is NOW! Data is information, and information is power."

~ **Radi,**
Data Analyst at CENTOGENE

SCADS
Powered by StarCore

Data Analytics?

Analytics is the use of:

data,

information technology,

statistical analysis,

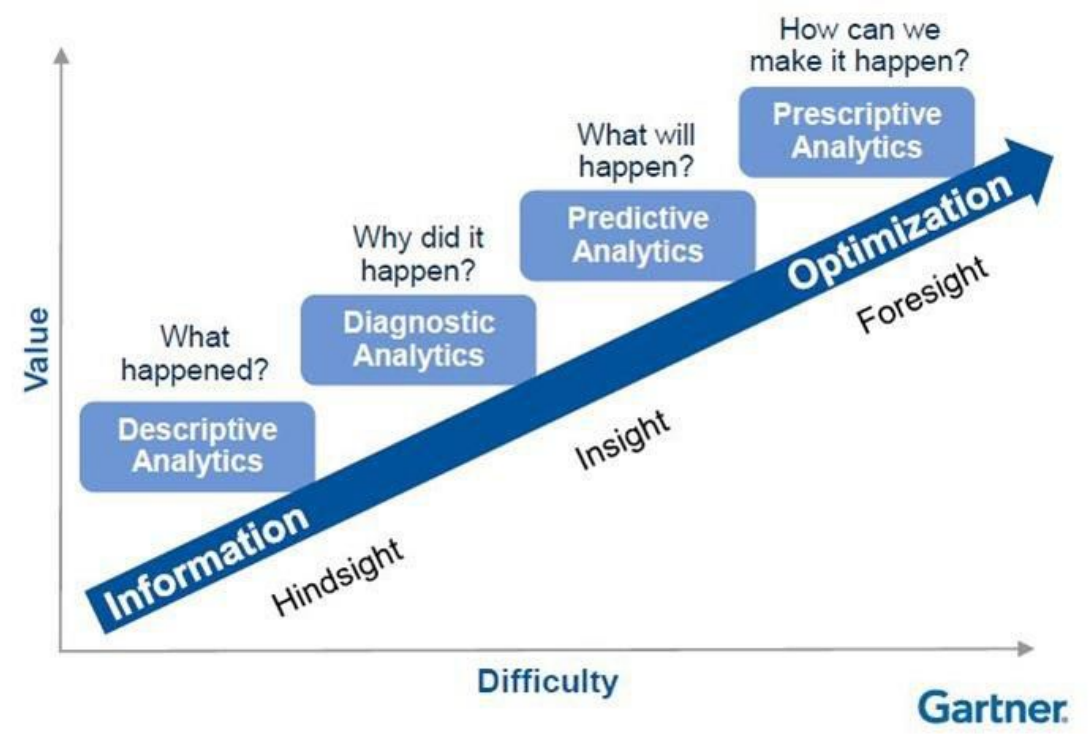
quantitative/qualitative methods, and

mathematical or computer-based models

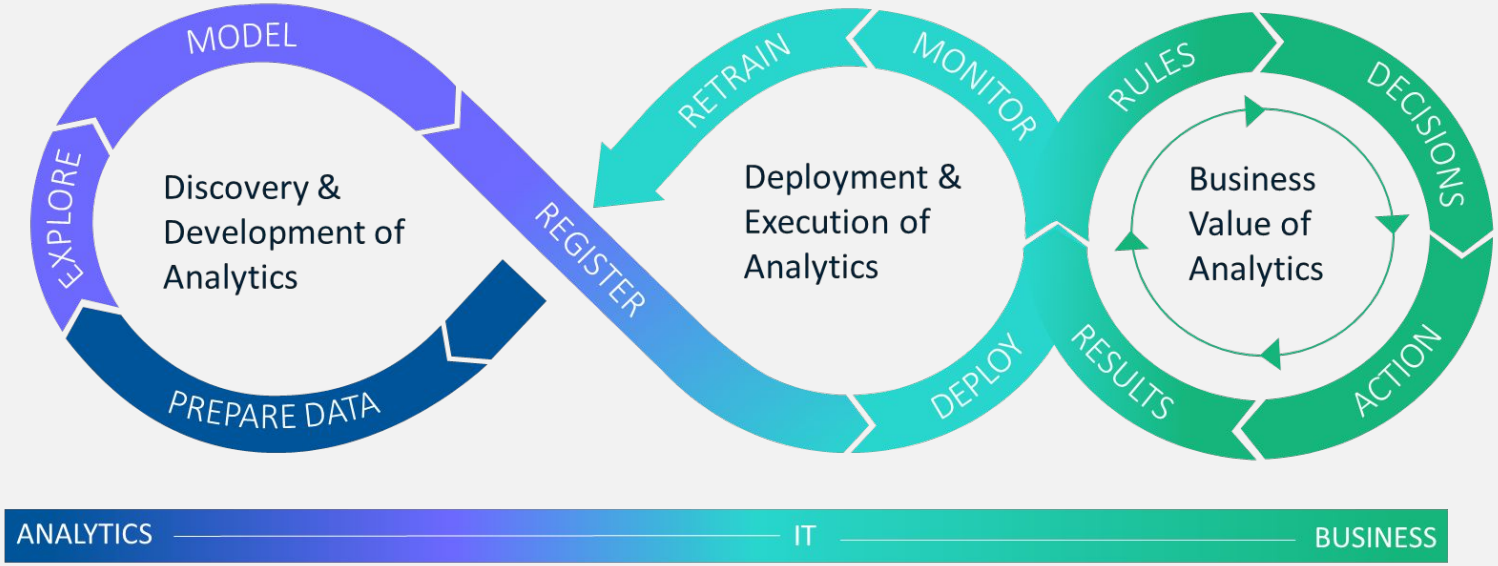
to help manager/decision maker gain improved insight about their business operations and make better, fact-based decisions

SCADS
Powered by StarCore

Data Analytics?



Data Analytics Lifecycle



SCADS
Powered by StarCore

Importance of Data Analytics

Enhances Data Understanding

Enable Creation of Informative Reports

to Remain Competitive Business

There is a strong relationship of DA with:

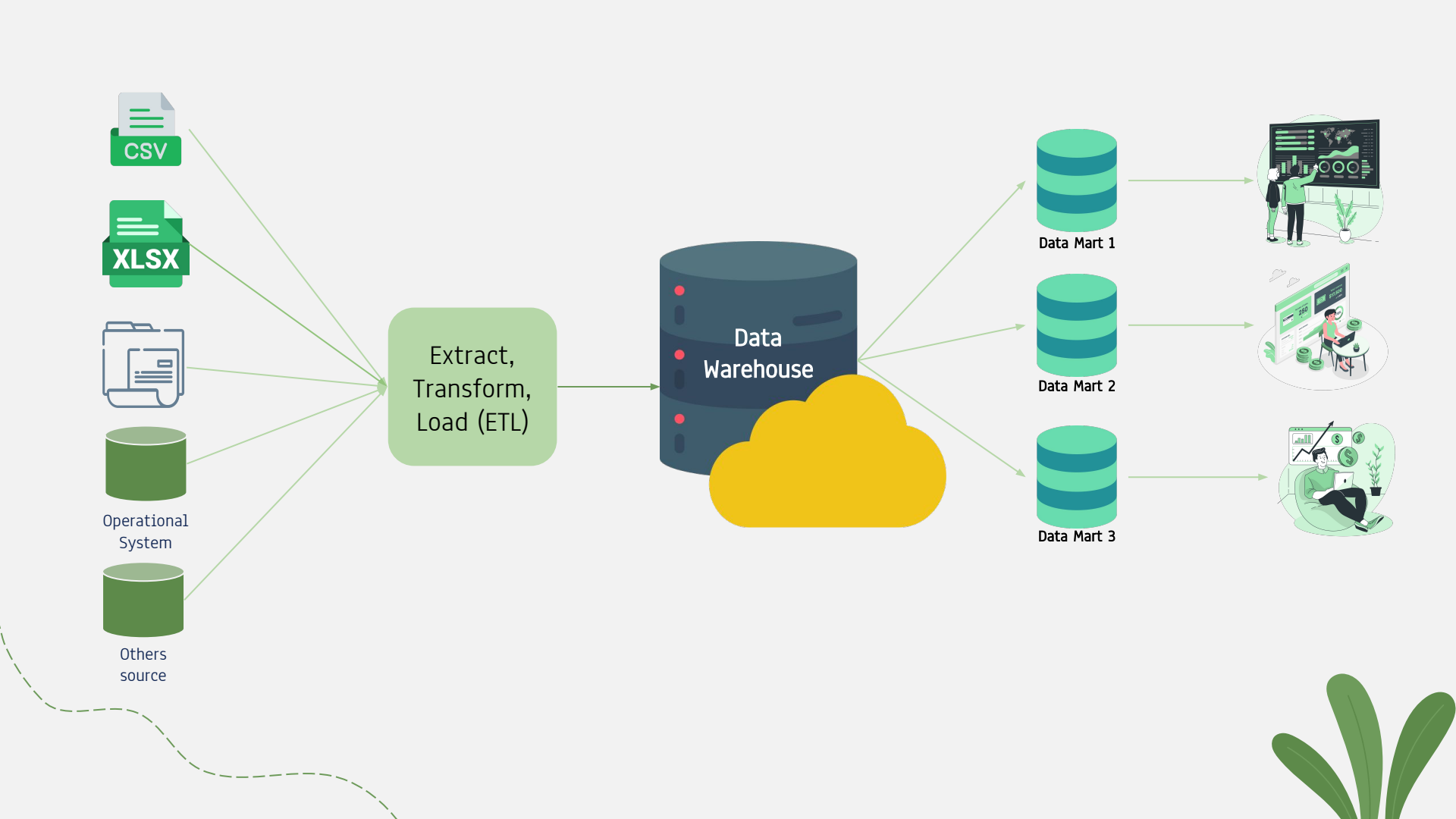
- profitability of businesses
- revenue of businesses
- shareholder return

Customer Segmentation

SCADS
Powered by StarCore

03

Data Mart for Customer Segmentation


CSV
XLSX
Operational System
Others source
Extract, Transform, Load (ETL)
Data Warehouse
Data Mart 1
Data Mart 2
Data Mart 3

SCADS
Powered by StarCore

Business Understanding

Company XYZ produces and sells baby products, one of which is baby milk. You can purchase baby milk with a membership or not.

Purchasing with membership allows customers to take part in loyalty programs, and also makes it possible for companies to carry out Customer Segmentation using Data Science on customers with membership.

SCADS
Powered by StarCore

About Datasets

The data is record of baby milk purchases by each customer with membership. The transaction period taken for 12 months (January to December 2019) was withdrawn on **01 January 2020**.

Member

Member Master Data

Transaction

Transaction Record Master Data

Subproduct

Product Detail Master Data

SCADS
Powered by StarCore

Member

Member Master Data

Transaction

Transaction Record Master Data

Subproduct

Product Detail Master Data

A	B	C	D
MemberID	JoinDate	NoChild	YoungestKidDOB
308735	2018-04-23	1	2016-10-15
338053	2019-07-09	1	2017-10-27
380615	2018-10-30	1	2016-09-28
430272	2019-09-04	1	2018-07-17
126108	2017-07-20	1	2015-11-24
364730	2019-07-02	1	2018-06-10
376959	2018-12-01	1	2015-06-04
70792	2017-05-16	1	2015-10-17
98552	2018-10-10	1	2017-06-17
213808	2018-12-24	1	2017-06-19
101762	2019-10-31	1	2018-03-29
309142	2015-12-10	1	2014-04-20

Transaction

Transaction Record Master Data

Subproduct

Product Detail Master Data

Member

Master Member Data

	A	B	C	D	E	F	G
1	MemberID	FKProductID	FKSubProductID	Qty	PricePerUnit	TransactionDate	TransactionTime
2	32269	27	52	1	139	2019-01-01	11:02:00
3	219312	24	67	1	383	2019-01-01	14:37:00
4	198412	27	52	3	156	2019-01-01	14:43:00
5	198412	26	49	3	80	2019-01-01	14:43:00
6	308087	26	50	1	134	2019-01-01	14:46:00
7	349520	26	49	1	82	2019-01-01	18:13:00
8	357289	22	39	1	244	2019-01-01	18:39:00
9	357289	27	52	1	139	2019-01-01	18:39:00
10	357289	27	69	1	258	2019-01-01	18:39:00
11	390918	26	50	1	128	2019-01-01	19:33:00
12	146715	26	50	1	125	2019-01-01	19:34:00
13	304517	27	69	1	260	2019-01-01	21:09:00
14	115868	27	51	1	87	2019-01-01	21:39:00
15	83600	24	67	2	385	2019-01-02	11:41:00
16	291772	26	49	2	79	2019-01-02	15:46:00
17	120348	24	45	1	221	2019-01-02	17:36:00

SCADS
Powered by StarCore

Subproduct

Product Detail Master Data

Member

Master Member Data

Transaction

Transaction Record Master Data

	A	B	C
1	SubProductID	Weight	Point
2	1	4	2
3	2	4	2
4	3	4	5
5	4	4	5
6	5	9	10
7	6	9	10
8	7	9	44
9	8	9	69
10	9	9	44
11	10	9	69
12	11	9	10
13	12	9	10
14	13	7	5

Analytical Base Table

	A	B	C	D	E	F	G	H	I	J	K	L
1	MemberID	NoChild	tenure_months	youngest_kid_age_join	recency	frequency	monetary	avg_monthly_frequency	avg_monetary	freq_last3mo	avg_interpurchase	avg_consumption
2	32269	1	12.91	1.08	15.50	43	13951	3.58	324.44	6	8.31	14.30
3	219312	2	28.75	2.48	1.39	18	10576	1.80	587.56	5	21.35	24.00
4	198412	3	22.57	1.33	14.28	60	22409	5.00	373.48	14	5.93	17.70
5	308087	1	22.90	4.02	332.60	4	729	2.00	182.25	0	10.67	9.75
6	349520	1	36.83	2.55	10.60	69	5668	5.75	82.14	21	5.21	4.00
7	357289	2	31.77	1.03	325.29	2	1076	1.00	538.00	0	39.00	25.50
8	390918	2	26.84	2.28	144.21	4	994	1.33	248.50	0	73.33	9.25
9	146715	1	13.50	3.03	345.17	3	375	3.00	125.00	0	9.50	7.00
10	304517	1	16.79	1.71	248.43	7	822	2.33	117.43	0	19.33	5.43
11	115868	1	19.38	1.09	223.18	45	4245	9.00	94.33	0	3.20	4.13
12	83600	3	32.79	2.54	29.54	10	5972	1.11	597.20	2	37.11	24.20
13	291772	2	36.34	1.48	62.24	11	2012	1.83	182.91	1	30.10	9.73
14	120348	2	15.15	1.74	269.12	4	894	1.33	223.50	0	31.33	9.00
15	96074	2	31.05	2.55	26.47	20	3146	2.22	157.30	6	17.74	8.10
16	399805	1	33.38	1.05	5.28	7	2390	1.17	341.43	1	59.50	17.57
17	234043	1	30.98	1.26	247.58	39	3725	9.75	95.51	0	3.03	4.33
18	195634	1	13.73	2.84	14.34	7	3814	1.17	544.86	1	58.00	26.43
19	78823	1	27.20	1.58	168.31	12	2847	2.00	237.25	0	17.64	9.00
20	149907	1	14.06	1.66	9.19	51	14536	4.64	285.02	15	7.06	9.00
21	120272	1	15.44	1.44	0.31	24	5382	2.40	224.25	8	15.70	11.04

SCADS
Powered by StarCore

04

Hands-on Data Preparation and Exploration

SCADS
Powered by StarCore

Day 2

SCADS
Powered by StarCore

Table of contents

SCADS
Powered by StarCore

01

Introduction to Cluster Analysis

SCADS
Powered by StarCore

Distance as Core of Cluster Analysis

- Distance is often equated with the term “dissimilarity”
- Suppose there are two individuals/objects, A and B, the distance between both denoted with d(A, B). The property of distance is:
  - $d(A,B) = d(B,A) \geq 0$
  - $d(A,A) = d(B,B) = 0$
  - If there is another individual, C, then $d(A,B) \leq d(A,C) + d(C,B)$
- There are various formulas/definitions for distance measure:
  - **Euclidean**, Weighted Euclidean, Mahalanobis, City block, …
  - Jaccard, Hamming, …

SCADS
Powered by StarCore

Euclidean Distance

$$d(A,B) = \sqrt{\sum_{i=1}^{p}(a_i - b_i)^2}$$

p is number of attributes/variables,
a_i and b_i is value of object/individu A and B on i attribute/variable.

Name	Height	Weight	Age
Aep	158	80	32
Iqbal	170	70	40
Lia	170	55	45

SCADS
Powered by StarCore

Euclidean Distance

$$d(A,B) = \sqrt{\sum_{i=1}^{p}(a_i - b_i)^2}$$

Name	Height	Weight	Age
Aep	158	80	32
Iqbal	170	70	40
Lia	170	55	45

$$d(\text{Aep, Iqbal}) = \sqrt{(158-170)^2 + (80-70)^2 + (32-40)^2} = 17.55$$

$$d(\text{Aep, Lia}) = \sqrt{(158-170)^2 + (80-55)^2 + (32-45)^2} = 30.63$$

$$d(\text{Iqbal, Lia}) = \sqrt{(170-170)^2 + (70-55)^2 + (40-45)^2} = 15.81$$

Distance	Aep	Iqbal	Lia
Aep	0	17.55	30.63
Iqbal	17.55	0	15. 81
Lia	30.63	15.81	0

SCADS
Powered by StarCore

Euclidean Distance

Name	Height (cm)	Weight	Age
Aep	158	80	32
Iqbal	170	70	40
Lia	170	55	45

Name	Height (m)	Weight	Age
Aep	1.58	80	32
Iqbal	1.70	70	40
Lia	1.70	55	45

Distance	Aep	Iqbal	Lia
Aep	0	17.55	30.63
Iqbal	17.55	0	15. 81
Lia	30.63	15.81	0

Distance	Aep	Iqbal	Lia
Aep	0	12.81	28.18
Iqbal	12.81	0	15.81
Lia	28.18	15.81	0

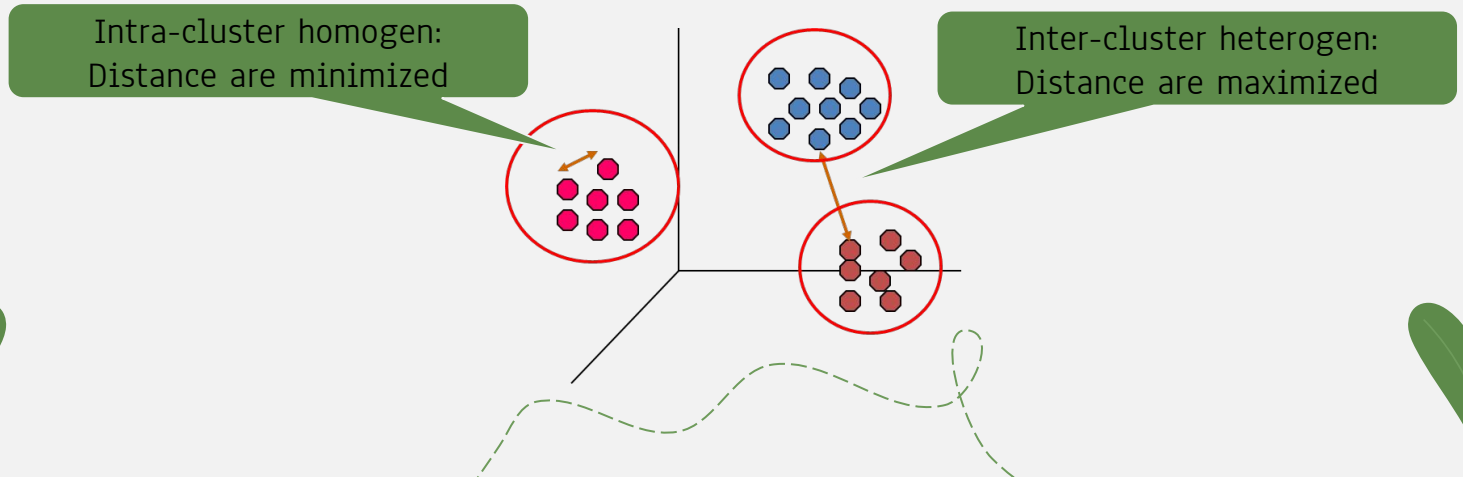
SCADS
Powered by StarCore

02

Concept of Cluster Analysis using K-means

Cluster Analysis

- **Finding groups** of objects/individuals "**by nature**" **based on similarity** and dissimilarity of their characteristics.
- Objects in a **same group** (intra-cluster) have **similar characteristics** (homogen) while objects from **different groups** (inter-cluster) have **different characteristics** (heterogen).

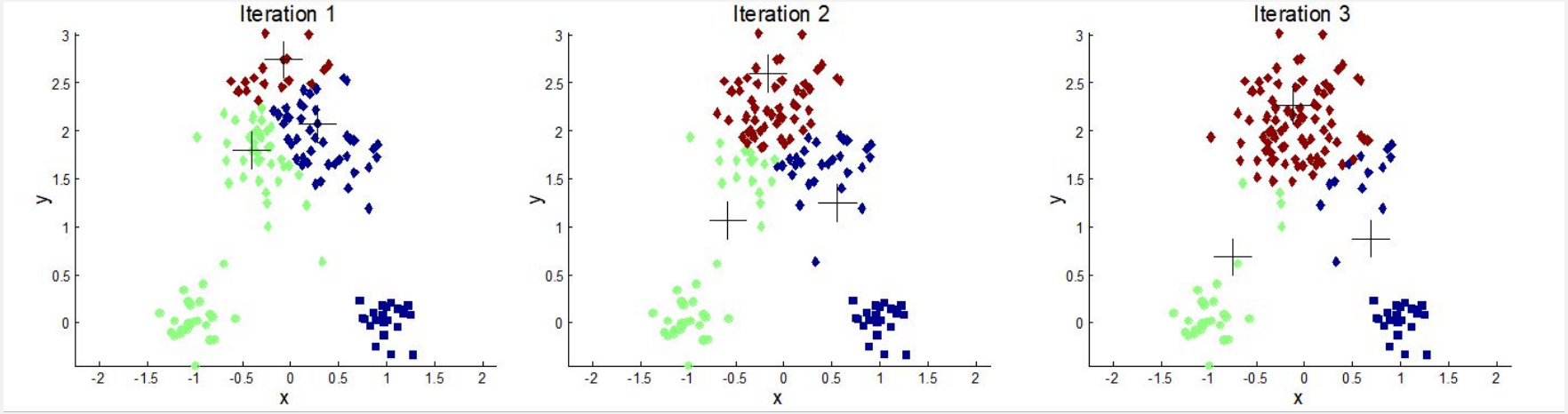


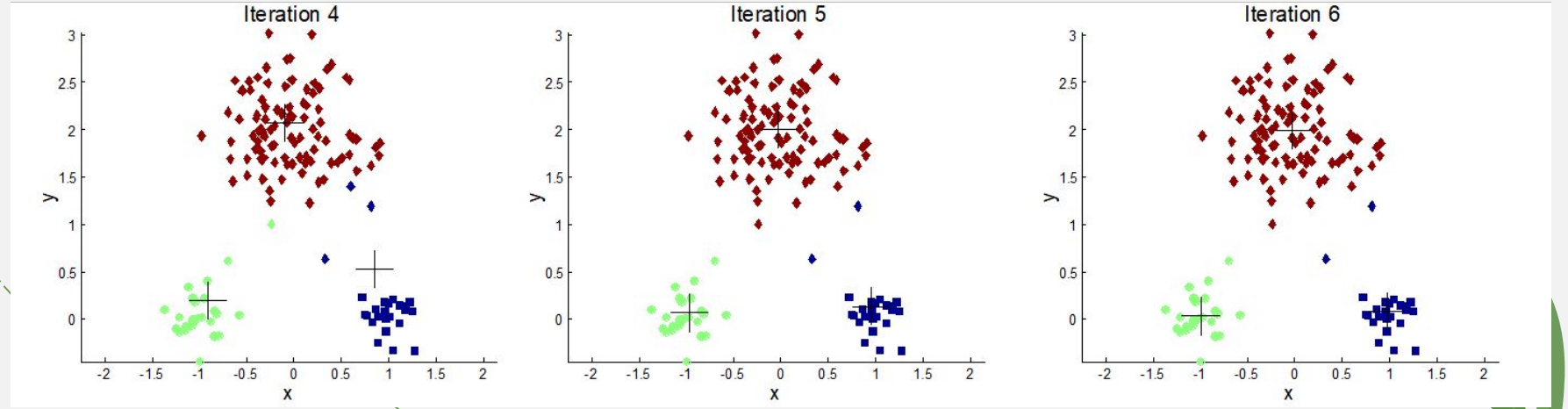
Objective Cluster Analysis

ID	X1	X2	X3	Cluster
1	1.5	80	15	2
2	1	102.5	7	1
3	2.3	56	10	3
4	1.8	92	6	1
⋮	⋮	⋮	⋮	⋮

K-means Algorithm

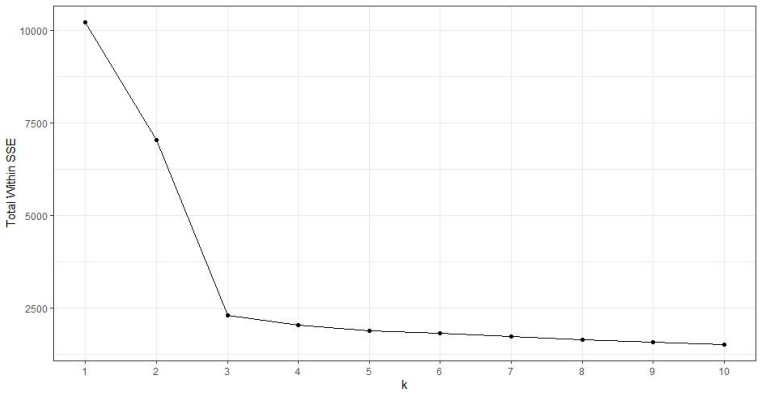
1. The number ***k*** of clusters is fixed
2. An initial set of ***k*** "seeds" (aggregation centres) is provided
    - First *k* elements
    - Other seeds (randomly selected or explicitly defined)
3. Given a certain fixed threshold, all units are assigned to the nearest cluster seed
4. New seeds are computed
5. Go back to step 3 until no reclassification is necessary
6. Units can be reassigned in successive steps (optimising partitioning)

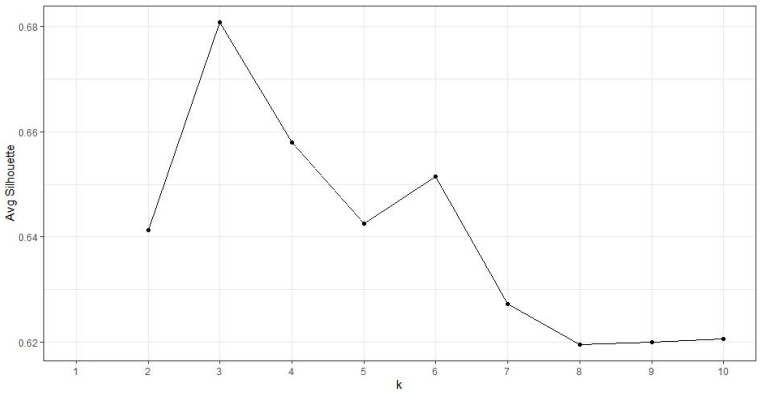

Iteration 1
Iteration 2
Iteration 3
x
y


Iteration 4
Iteration 5
Iteration 6
x
y

Number of *K*

- Approach:
  - Determined by the analysis or theoretical
  - Fixed by the researchers
- Preferable approach: "*let the data speak*"

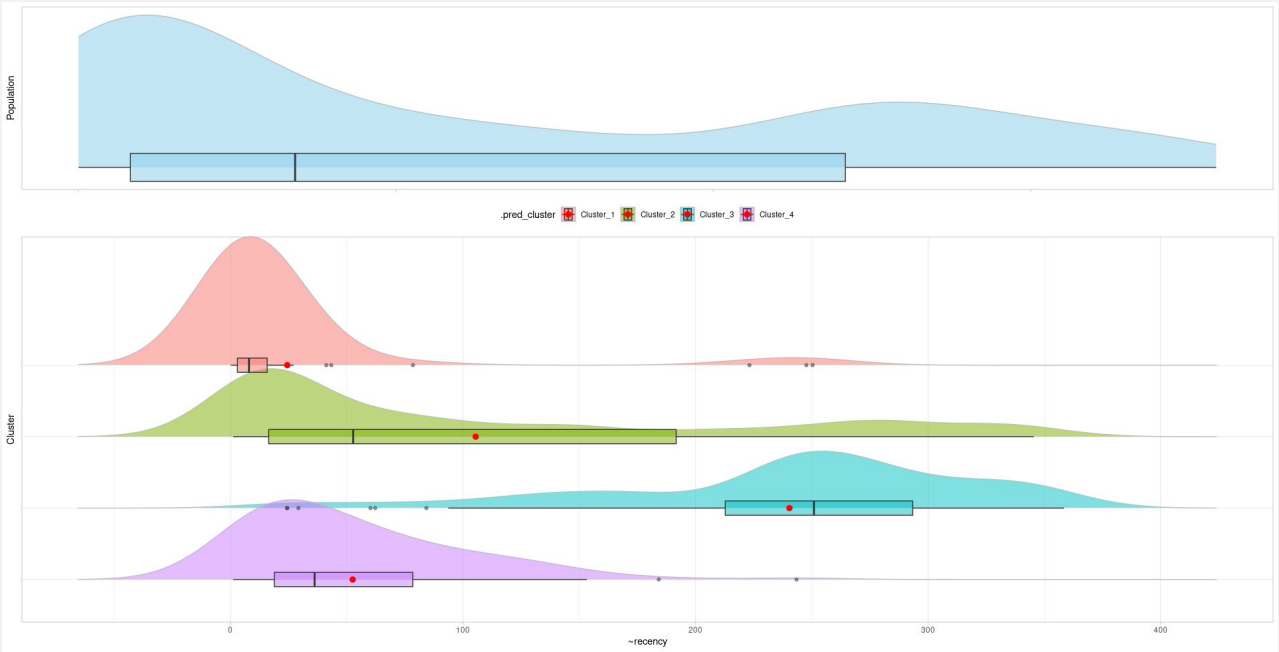




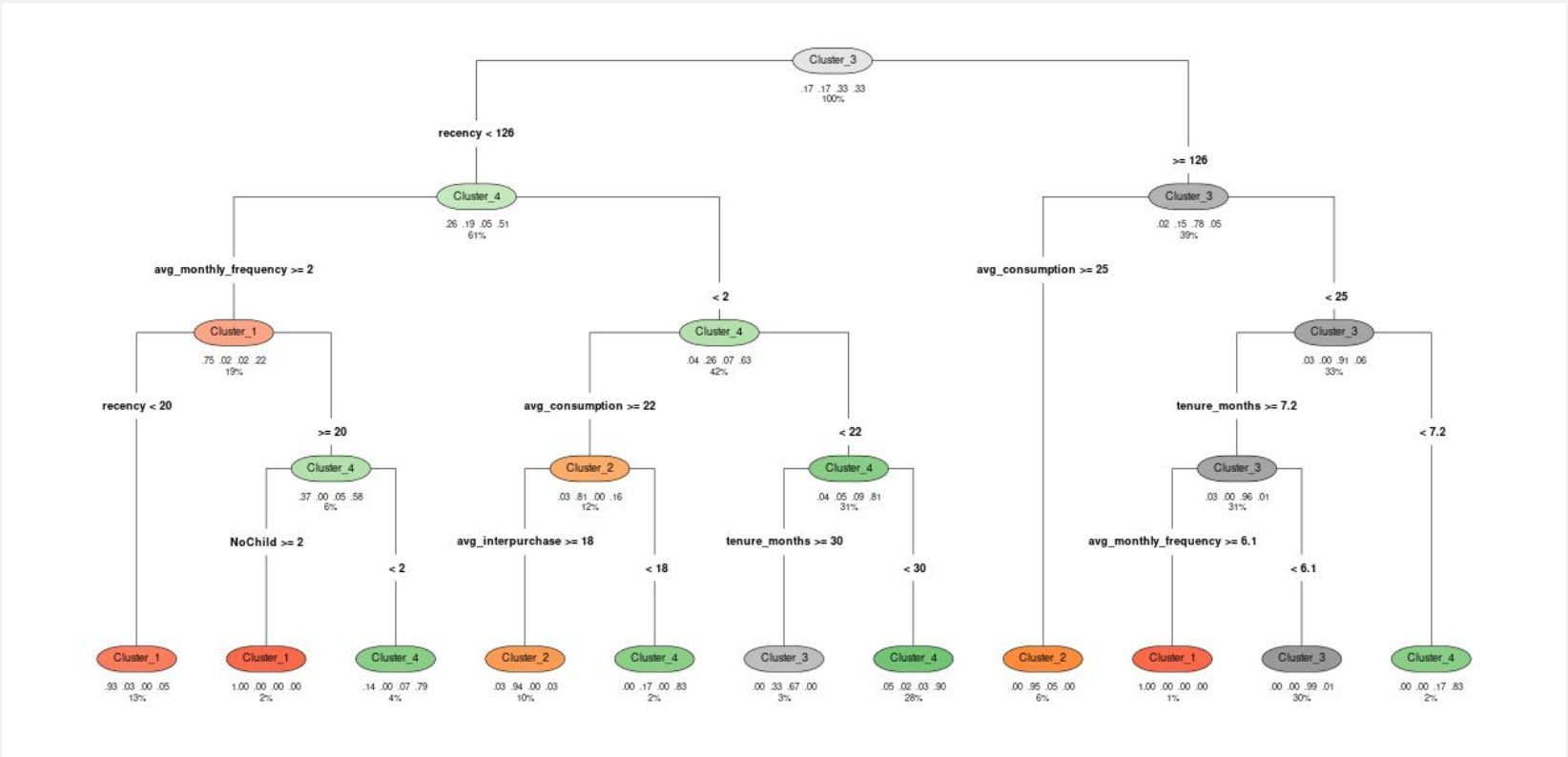
Profiling Segment

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Recency	24.4	105	240	52.5
Avg Monthly Monetary	280	857	247	295
Avg Monthly Frequency	3.35	1.31	1.55	1.51
Tenure	11.8	22.9	20.8	8.04

Profiling Segment



Profiling Segment



SCADS
Powered by StarCore

03 Hands-on Customer Segmentation

Analytical Base Table

	A	B	C	D	E	F	G	H	I	J	K	L
1	MemberID	NoChild	tenure_months	youngest_kid_age_join	recency	frequency	monetary	avg_monthly_frequency	avg_monetary	freq_last3mo	avg_interpurchase	avg_consumption
2	32269	1	12.91	1.08	15.50	43	13951	3.58	324.44	6	8.31	14.30
3	219312	2	28.75	2.48	1.39	18	10576	1.80	587.56	5	21.35	24.00
4	198412	3	22.57	1.33	14.28	60	22409	5.00	373.48	14	5.93	17.70
5	308087	1	22.90	4.02	332.60	4	729	2.00	182.25	0	10.67	9.75
6	349520	1	36.83	2.55	10.60	69	5668	5.75	82.14	21	5.21	4.00
7	357289	2	31.77	1.03	325.29	2	1076	1.00	538.00	0	39.00	25.50
8	390918	2	26.84	2.28	144.21	4	994	1.33	248.50	0	73.33	9.25
9	146715	1	13.50	3.03	345.17	3	375	3.00	125.00	0	9.50	7.00
10	304517	1	16.79	1.71	248.43	7	822	2.33	117.43	0	19.33	5.43
11	115868	1	19.38	1.09	223.18	45	4245	9.00	94.33	0	3.20	4.13
12	83600	3	32.79	2.54	29.54	10	5972	1.11	597.20	2	37.11	24.20
13	291772	2	36.34	1.48	62.24	11	2012	1.83	182.91	1	30.10	9.73
14	120348	2	15.15	1.74	269.12	4	894	1.33	223.50	0	31.33	9.00
15	96074	2	31.05	2.55	26.47	20	3146	2.22	157.30	6	17.74	8.10
16	399805	1	33.38	1.05	5.28	7	2390	1.17	341.43	1	59.50	17.57
17	234043	1	30.98	1.26	247.58	39	3725	9.75	95.51	0	3.03	4.33
18	195634	1	13.73	2.84	14.34	7	3814	1.17	544.86	1	58.00	26.43
19	78823	1	27.20	1.58	168.31	12	2847	2.00	237.25	0	17.64	9.00
20	149907	1	14.06	1.66	9.19	51	14536	4.64	285.02	15	7.06	9.00
21	120272	1	15.44	1.44	0.31	24	5382	2.40	224.25	8	15.70	11.04

About Dataset

MemberID	ID pelanggan bersifat unik
NoChild	No of Child
tenure_months	Length of stay as member (month)
youngest_kid_age_join	Youngest kid's age when join membership (year)
recency	Length of the last transaction from the date of data withdrawal (days)
frequency	Number of transaction within last 12 months
monetary	Total amount spent (x1000 rupiah)
avg_monthly_frequency	Average number of transaction per month (only on month which there are transaction)
avg_monetary	Average amount spent each transaction (x1000 rupiah)
avg_consumption	Average weight of product purchases per transaction (ounces)
freq_last3mo	Number of transactions in the last 3 months
avg_interpurchase	Average time distance between a transaction and the next transaction (days); (0 = no next transaction / continues shopping every day)

Thanks!

Do you have any questions?

aep.hidayatuloh@starcore.co
https://starcore.co/scads