

customer-segmentation

Sofyan Adi Saputra

2024-01-29

1. DATA PREPARATION

Dataset yang digunakan : <https://www.kaggle.com/datasets/sofyanadi/customer-segmentation>

```
#Load Dataset
#path/lokasi dataset
df <- read.csv("C:/Users/sofyan/Documents/R/Mini Project/customer-segmentation/Datasets/customer_segmen

#Vector untuk Menyimpan Nama Field
field_yang_digunakan <- df[c("Jenis.Kelamin", "Umur", "Profesi")]
field_yang_digunakan
```

##	Jenis.Kelamin	Umur	Profesi
## 1	Pria	58	Wiraswasta
## 2	Wanita	14	Pelajar
## 3	Pria	48	Professional
## 4	Pria	53	Professional
## 5	Wanita	41	Wiraswasta
## 6	Wanita	24	Professional
## 7	Pria	64	Wiraswasta
## 8	Pria	52	Professional
## 9	Wanita	29	Professional
## 10	Pria	33	Professional
## 11	Wanita	50	Professional
## 12	Wanita	49	Professional
## 13	Wanita	64	Wiraswasta
## 14	Pria	60	Wiraswasta
## 15	Wanita	20	Wiraswasta
## 16	Pria	35	Professional
## 17	Wanita	32	Ibu Rumah Tangga
## 18	Wanita	63	Ibu Rumah Tangga
## 19	Wanita	32	Wiraswasta
## 20	Wanita	16	Pelajar
## 21	Wanita	38	Wiraswasta
## 22	Wanita	52	Professional
## 23	Pria	34	Professional
## 24	Wanita	39	Wiraswasta
## 25	Wanita	29	Wiraswasta
## 26	Wanita	55	Professional
## 27	Wanita	35	Wiraswasta
## 28	Wanita	40	Ibu Rumah Tangga

```
## 29      Wanita  56      Professional
## 30      Wanita  46 Ibu Rumah Tangga
## 31      Wanita  19      Mahasiswa
## 32      Wanita  47      Wiraswasta
## 33      Wanita  19      Mahasiswa
## 34      Wanita  21      Wiraswasta
## 35      Wanita  39      Professional
## 36      Wanita  30      Wiraswasta
## 37      Wanita  25      Professional
## 38      Wanita  46      Wiraswasta
## 39      Wanita  20      Professional
## 40      Wanita  14      Pelajar
## 41      Wanita  24 Ibu Rumah Tangga
## 42      Wanita  26      Wiraswasta
## 43      Wanita  31      Professional
## 44      Wanita  18      Wiraswasta
## 45      Wanita  22      Professional
## 46      Wanita  25      Wiraswasta
## 47      Wanita  55 Ibu Rumah Tangga
## 48      Wanita  45      Wiraswasta
## 49      Wanita  33 Ibu Rumah Tangga
## 50      Wanita  55      Wiraswasta
```

#Konversi Data dengan data.matrix

```
pelanggan_matrix <- data.matrix(df[c("Jenis.Kelamin", "Profesi", "Tipe.Residen")])
pelanggan_matrix
```

```
##      Jenis.Kelamin Profesi Tipe.Residen
## [1,]              1      5             2
## [2,]              2      3             1
## [3,]              1      4             1
## [4,]              1      4             1
## [5,]              2      5             1
## [6,]              2      4             1
## [7,]              1      5             2
## [8,]              1      4             1
## [9,]              2      4             2
## [10,]             1      4             1
## [11,]             2      4             2
## [12,]             2      4             2
## [13,]             2      5             1
## [14,]             1      5             1
## [15,]             2      5             1
## [16,]             1      4             1
## [17,]             2      1             1
## [18,]             2      1             1
## [19,]             2      5             1
## [20,]             2      3             2
## [21,]             2      5             1
## [22,]             2      4             1
## [23,]             1      4             1
## [24,]             2      5             1
## [25,]             2      5             2
## [26,]             2      4             1
```

```
## [27,]      2      5      1
## [28,]      2      1      1
## [29,]      2      4      1
## [30,]      2      1      2
## [31,]      2      2      1
## [32,]      2      5      2
## [33,]      2      2      1
## [34,]      2      5      2
## [35,]      2      4      2
## [36,]      2      5      1
## [37,]      2      4      2
## [38,]      2      5      2
## [39,]      2      4      1
## [40,]      2      3      2
## [41,]      2      1      1
## [42,]      2      5      1
## [43,]      2      4      1
## [44,]      2      5      1
## [45,]      2      4      1
## [46,]      2      5      2
## [47,]      2      1      1
## [48,]      2      5      2
## [49,]      2      1      2
## [50,]      2      5      2
```

```
#Penggabungan data pelanggan dan data pelanggan_matriks
pelanggan <- data.frame(df, pelanggan_matrix)
pelanggan
```

```
##      Customer_ID      Nama.Pelanggan Jenis.Kelamin Umur      Profesi
## 1      CUST-001      Budi Anggara      Pria      58      Wiraswasta
## 2      CUST-002      Shirley Ratuwati      Wanita      14      Pelajar
## 3      CUST-003      Agus Cahyono      Pria      48      Professional
## 4      CUST-004      Antonius Winarta      Pria      53      Professional
## 5      CUST-005      Ibu Sri Wahyuni, IR      Wanita      41      Wiraswasta
## 6      CUST-006      Rosalina Kurnia      Wanita      24      Professional
## 7      CUST-007      Cahyono, Agus      Pria      64      Wiraswasta
## 8      CUST-008      Danang Santosa      Pria      52      Professional
## 9      CUST-009      Elisabeth Suryadinata      Wanita      29      Professional
## 10     CUST-010      Mario Setiawan      Pria      33      Professional
## 11     CUST-011      Maria Suryawan      Wanita      50      Professional
## 12     CUST-012      Erliana Widjaja      Wanita      49      Professional
## 13     CUST-013      Cahaya Putri      Wanita      64      Wiraswasta
## 14     CUST-014      Mario Setiawan      Pria      60      Wiraswasta
## 15     CUST-015      Shirley Ratuwati      Wanita      20      Wiraswasta
## 16     CUST-016      Bambang Rudi      Pria      35      Professional
## 17     CUST-017      Yuni Sari      Wanita      32      Ibu Rumah Tangga
## 18     CUST-018      Nelly Halim      Wanita      63      Ibu Rumah Tangga
## 19     CUST-019      Mega Pranoto      Wanita      32      Wiraswasta
## 20     CUST-020      Irene Novianto      Wanita      16      Pelajar
## 21     CUST-021      Lestari Fabianto      Wanita      38      Wiraswasta
## 22     CUST-022      Novita Purba      Wanita      52      Professional
## 23     CUST-023      Denny Amiruddin      Pria      34      Professional
## 24     CUST-024      Putri Ginting      Wanita      39      Wiraswasta
```

## 25	CUST-025	Julia Setiawan	Wanita	29	Wiraswasta
## 26	CUST-026	Christine Winarto	Wanita	55	Professional
## 27	CUST-027	Grace Mulyati	Wanita	35	Wiraswasta
## 28	CUST-028	Adeline Huang	Wanita	40	Ibu Rumah Tangga
## 29	CUST-029	Tia Hartanti	Wanita	56	Professional
## 30	CUST-030	Rosita Saragih	Wanita	46	Ibu Rumah Tangga
## 31	CUST-031	Eviana Handry	Wanita	19	Mahasiswa
## 32	CUST-032	Chintya Winarni	Wanita	47	Wiraswasta
## 33	CUST-033	Cecilia Kusnadi	Wanita	19	Mahasiswa
## 34	CUST-034	Deasy Arisandi	Wanita	21	Wiraswasta
## 35	CUST-035	Ida Ayu	Wanita	39	Professional
## 36	CUST-036	Ni Made Suasti	Wanita	30	Wiraswasta
## 37	CUST-037	Felicia Tandiono	Wanita	25	Professional
## 38	CUST-038	Agatha Salim	Wanita	46	Wiraswasta
## 39	CUST-039	Gina Hidayat	Wanita	20	Professional
## 40	CUST-040	Irene Darmawan	Wanita	14	Pelajar
## 41	CUST-041	Shinta Aritonang	Wanita	24	Ibu Rumah Tangga
## 42	CUST-042	Yuliana Wati	Wanita	26	Wiraswasta
## 43	CUST-043	Yenna Sumadi	Wanita	31	Professional
## 44	CUST-044	Anna	Wanita	18	Wiraswasta
## 45	CUST-045	Rismawati Juni	Wanita	22	Professional
## 46	CUST-046	Elfira Surya	Wanita	25	Wiraswasta
## 47	CUST-047	Mira Kurnia	Wanita	55	Ibu Rumah Tangga
## 48	CUST-048	Maria Hutagalung	Wanita	45	Wiraswasta
## 49	CUST-049	Josephine Wahab	Wanita	33	Ibu Rumah Tangga
## 50	CUST-050	Lianna Nugraha	Wanita	55	Wiraswasta
##	Tipe.Residen	NilaiBelanjaSetahun	Jenis.Kelamin.1	Profesi.1	Tipe.Residen.1
## 1	Sector	9497927	1	5	2
## 2	Cluster	2722700	2	3	1
## 3	Cluster	5286429	1	4	1
## 4	Cluster	5204498	1	4	1
## 5	Cluster	10615206	2	5	1
## 6	Cluster	5215541	2	4	1
## 7	Sector	9837260	1	5	2
## 8	Cluster	5223569	1	4	1
## 9	Sector	5993218	2	4	2
## 10	Cluster	5257448	1	4	1
## 11	Sector	5987367	2	4	2
## 12	Sector	5941914	2	4	2
## 13	Cluster	9333168	2	5	1
## 14	Cluster	9471615	1	5	1
## 15	Cluster	10365668	2	5	1
## 16	Cluster	5262521	1	4	1
## 17	Cluster	5677762	2	1	1
## 18	Cluster	5340690	2	1	1
## 19	Cluster	10884508	2	5	1
## 20	Sector	2896845	2	3	2
## 21	Cluster	9222070	2	5	1
## 22	Cluster	5298157	2	4	1
## 23	Cluster	5239290	1	4	1
## 24	Cluster	10259572	2	5	1
## 25	Sector	10721998	2	5	2
## 26	Cluster	5269392	2	4	1
## 27	Cluster	9114159	2	5	1

## 28	Cluster	6631680	2	1	1
## 29	Cluster	5271845	2	4	1
## 30	Sector	5020976	2	1	2
## 31	Cluster	3042773	2	2	1
## 32	Sector	10663179	2	5	2
## 33	Cluster	3047926	2	2	1
## 34	Sector	9759822	2	5	2
## 35	Sector	5962575	2	4	2
## 36	Cluster	9678994	2	5	1
## 37	Sector	5972787	2	4	2
## 38	Sector	10477127	2	5	2
## 39	Cluster	5257775	2	4	1
## 40	Sector	2861855	2	3	2
## 41	Cluster	6820976	2	1	1
## 42	Cluster	9880607	2	5	1
## 43	Cluster	5268410	2	4	1
## 44	Cluster	9339737	2	5	1
## 45	Cluster	5211041	2	4	1
## 46	Sector	10099807	2	5	2
## 47	Cluster	6130724	2	1	1
## 48	Sector	10390732	2	5	2
## 49	Sector	4992585	2	1	2
## 50	Sector	10569316	2	5	2

#Menormalisasikan Nilai Belanja

```
pelanggan$NilaiBelanjaSetahun <- pelanggan$NilaiBelanjaSetahun / 1000000
pelanggan
```

##	Customer_ID	Nama.Pelanggan	Jenis.Kelamin	Umur	Profesi
## 1	CUST-001	Budi Anggara	Pria	58	Wiraswasta
## 2	CUST-002	Shirley Ratuwati	Wanita	14	Pelajar
## 3	CUST-003	Agus Cahyono	Pria	48	Professional
## 4	CUST-004	Antonius Winarta	Pria	53	Professional
## 5	CUST-005	Ibu Sri Wahyuni, IR	Wanita	41	Wiraswasta
## 6	CUST-006	Rosalina Kurnia	Wanita	24	Professional
## 7	CUST-007	Cahyono, Agus	Pria	64	Wiraswasta
## 8	CUST-008	Danang Santosa	Pria	52	Professional
## 9	CUST-009	Elisabeth Suryadinata	Wanita	29	Professional
## 10	CUST-010	Mario Setiawan	Pria	33	Professional
## 11	CUST-011	Maria Suryawan	Wanita	50	Professional
## 12	CUST-012	Erliana Widjaja	Wanita	49	Professional
## 13	CUST-013	Cahaya Putri	Wanita	64	Wiraswasta
## 14	CUST-014	Mario Setiawan	Pria	60	Wiraswasta
## 15	CUST-015	Shirley Ratuwati	Wanita	20	Wiraswasta
## 16	CUST-016	Bambang Rudi	Pria	35	Professional
## 17	CUST-017	Yuni Sari	Wanita	32	Ibu Rumah Tangga
## 18	CUST-018	Nelly Halim	Wanita	63	Ibu Rumah Tangga
## 19	CUST-019	Mega Pranoto	Wanita	32	Wiraswasta
## 20	CUST-020	Irene Novianto	Wanita	16	Pelajar
## 21	CUST-021	Lestari Fabianto	Wanita	38	Wiraswasta
## 22	CUST-022	Novita Purba	Wanita	52	Professional
## 23	CUST-023	Denny Amiruddin	Pria	34	Professional
## 24	CUST-024	Putri Ginting	Wanita	39	Wiraswasta
## 25	CUST-025	Julia Setiawan	Wanita	29	Wiraswasta

## 26	CUST-026	Christine Winarto	Wanita	55	Professional
## 27	CUST-027	Grace Mulyati	Wanita	35	Wiraswasta
## 28	CUST-028	Adeline Huang	Wanita	40	Ibu Rumah Tangga
## 29	CUST-029	Tia Hartanti	Wanita	56	Professional
## 30	CUST-030	Rosita Saragih	Wanita	46	Ibu Rumah Tangga
## 31	CUST-031	Eviana Handry	Wanita	19	Mahasiswa
## 32	CUST-032	Chintya Winarni	Wanita	47	Wiraswasta
## 33	CUST-033	Cecilia Kusnadi	Wanita	19	Mahasiswa
## 34	CUST-034	Deasy Arisandi	Wanita	21	Wiraswasta
## 35	CUST-035	Ida Ayu	Wanita	39	Professional
## 36	CUST-036	Ni Made Suasti	Wanita	30	Wiraswasta
## 37	CUST-037	Felicia Tandiono	Wanita	25	Professional
## 38	CUST-038	Agatha Salim	Wanita	46	Wiraswasta
## 39	CUST-039	Gina Hidayat	Wanita	20	Professional
## 40	CUST-040	Irene Darmawan	Wanita	14	Pelajar
## 41	CUST-041	Shinta Aritonang	Wanita	24	Ibu Rumah Tangga
## 42	CUST-042	Yuliana Wati	Wanita	26	Wiraswasta
## 43	CUST-043	Yenna Sumadi	Wanita	31	Professional
## 44	CUST-044	Anna	Wanita	18	Wiraswasta
## 45	CUST-045	Rismawati Juni	Wanita	22	Professional
## 46	CUST-046	Elfira Surya	Wanita	25	Wiraswasta
## 47	CUST-047	Mira Kurnia	Wanita	55	Ibu Rumah Tangga
## 48	CUST-048	Maria Hutagalung	Wanita	45	Wiraswasta
## 49	CUST-049	Josephine Wahab	Wanita	33	Ibu Rumah Tangga
## 50	CUST-050	Lianna Nugraha	Wanita	55	Wiraswasta
##	Tipe.Residen	NilaiBelanjaSetahun	Jenis.Kelamin.1	Profesi.1	Tipe.Residen.1
## 1	Sector	9.497927	1	5	2
## 2	Cluster	2.722700	2	3	1
## 3	Cluster	5.286429	1	4	1
## 4	Cluster	5.204498	1	4	1
## 5	Cluster	10.615206	2	5	1
## 6	Cluster	5.215541	2	4	1
## 7	Sector	9.837260	1	5	2
## 8	Cluster	5.223569	1	4	1
## 9	Sector	5.993218	2	4	2
## 10	Cluster	5.257448	1	4	1
## 11	Sector	5.987367	2	4	2
## 12	Sector	5.941914	2	4	2
## 13	Cluster	9.333168	2	5	1
## 14	Cluster	9.471615	1	5	1
## 15	Cluster	10.365668	2	5	1
## 16	Cluster	5.262521	1	4	1
## 17	Cluster	5.677762	2	1	1
## 18	Cluster	5.340690	2	1	1
## 19	Cluster	10.884508	2	5	1
## 20	Sector	2.896845	2	3	2
## 21	Cluster	9.222070	2	5	1
## 22	Cluster	5.298157	2	4	1
## 23	Cluster	5.239290	1	4	1
## 24	Cluster	10.259572	2	5	1
## 25	Sector	10.721998	2	5	2
## 26	Cluster	5.269392	2	4	1
## 27	Cluster	9.114159	2	5	1
## 28	Cluster	6.631680	2	1	1

## 29	Cluster	5.271845	2	4	1
## 30	Sector	5.020976	2	1	2
## 31	Cluster	3.042773	2	2	1
## 32	Sector	10.663179	2	5	2
## 33	Cluster	3.047926	2	2	1
## 34	Sector	9.759822	2	5	2
## 35	Sector	5.962575	2	4	2
## 36	Cluster	9.678994	2	5	1
## 37	Sector	5.972787	2	4	2
## 38	Sector	10.477127	2	5	2
## 39	Cluster	5.257775	2	4	1
## 40	Sector	2.861855	2	3	2
## 41	Cluster	6.820976	2	1	1
## 42	Cluster	9.880607	2	5	1
## 43	Cluster	5.268410	2	4	1
## 44	Cluster	9.339737	2	5	1
## 45	Cluster	5.211041	2	4	1
## 46	Sector	10.099807	2	5	2
## 47	Cluster	6.130724	2	1	1
## 48	Sector	10.390732	2	5	2
## 49	Sector	4.992585	2	1	2
## 50	Sector	10.569316	2	5	2

#Membuat Data Master

```
Profesi <- unique(pelanggan[c("Profesi","Profesi.1")])
Jenis.Kelamin <- unique(pelanggan[c("Jenis.Kelamin","Jenis.Kelamin.1")])
Tipe.Residen <- unique(pelanggan[c("Tipe.Residen","Tipe.Residen.1")])
pelanggan[c("Profesi","Profesi.1", "Jenis.Kelamin", "Jenis.Kelamin.1","Tipe.Residen", "Tipe.Residen.1")]
```

##	Profesi	Profesi.1	Jenis.Kelamin	Jenis.Kelamin.1	Tipe.Residen
## 1	Wiraswasta	5	Pria	1	Sector
## 2	Pelajar	3	Wanita	2	Cluster
## 3	Professional	4	Pria	1	Cluster
## 4	Professional	4	Pria	1	Cluster
## 5	Wiraswasta	5	Wanita	2	Cluster
## 6	Professional	4	Wanita	2	Cluster
## 7	Wiraswasta	5	Pria	1	Sector
## 8	Professional	4	Pria	1	Cluster
## 9	Professional	4	Wanita	2	Sector
## 10	Professional	4	Pria	1	Cluster
## 11	Professional	4	Wanita	2	Sector
## 12	Professional	4	Wanita	2	Sector
## 13	Wiraswasta	5	Wanita	2	Cluster
## 14	Wiraswasta	5	Pria	1	Cluster
## 15	Wiraswasta	5	Wanita	2	Cluster
## 16	Professional	4	Pria	1	Cluster
## 17	Ibu Rumah Tangga	1	Wanita	2	Cluster
## 18	Ibu Rumah Tangga	1	Wanita	2	Cluster
## 19	Wiraswasta	5	Wanita	2	Cluster
## 20	Pelajar	3	Wanita	2	Sector
## 21	Wiraswasta	5	Wanita	2	Cluster
## 22	Professional	4	Wanita	2	Cluster
## 23	Professional	4	Pria	1	Cluster
## 24	Wiraswasta	5	Wanita	2	Cluster

## 25	Wiraswasta	5	Wanita	2	Sector
## 26	Professional	4	Wanita	2	Cluster
## 27	Wiraswasta	5	Wanita	2	Cluster
## 28	Ibu Rumah Tangga	1	Wanita	2	Cluster
## 29	Professional	4	Wanita	2	Cluster
## 30	Ibu Rumah Tangga	1	Wanita	2	Sector
## 31	Mahasiswa	2	Wanita	2	Cluster
## 32	Wiraswasta	5	Wanita	2	Sector
## 33	Mahasiswa	2	Wanita	2	Cluster
## 34	Wiraswasta	5	Wanita	2	Sector
## 35	Professional	4	Wanita	2	Sector
## 36	Wiraswasta	5	Wanita	2	Cluster
## 37	Professional	4	Wanita	2	Sector
## 38	Wiraswasta	5	Wanita	2	Sector
## 39	Professional	4	Wanita	2	Cluster
## 40	Pelajar	3	Wanita	2	Sector
## 41	Ibu Rumah Tangga	1	Wanita	2	Cluster
## 42	Wiraswasta	5	Wanita	2	Cluster
## 43	Professional	4	Wanita	2	Cluster
## 44	Wiraswasta	5	Wanita	2	Cluster
## 45	Professional	4	Wanita	2	Cluster
## 46	Wiraswasta	5	Wanita	2	Sector
## 47	Ibu Rumah Tangga	1	Wanita	2	Cluster
## 48	Wiraswasta	5	Wanita	2	Sector
## 49	Ibu Rumah Tangga	1	Wanita	2	Sector
## 50	Wiraswasta	5	Wanita	2	Sector
##	Tipe.Residen.1				
## 1	2				
## 2	1				
## 3	1				
## 4	1				
## 5	1				
## 6	1				
## 7	2				
## 8	1				
## 9	2				
## 10	1				
## 11	2				
## 12	2				
## 13	1				
## 14	1				
## 15	1				
## 16	1				
## 17	1				
## 18	1				
## 19	1				
## 20	2				
## 21	1				
## 22	1				
## 23	1				
## 24	1				
## 25	2				
## 26	1				
## 27	1				


```
## 28      1
## 29      1
## 30      2
## 31      1
## 32      2
## 33      1
## 34      2
## 35      2
## 36      1
## 37      2
## 38      2
## 39      1
## 40      2
## 41      1
## 42      1
## 43      1
## 44      1
## 45      1
## 46      2
## 47      1
## 48      2
## 49      2
## 50      2
```

Terlihat, teks “Pria” diubah menjadi angka 1, “Wanita” diubah menjadi angka 2, “Wiraswasta” pada kolom profesi diubah menjadi angka 5, “Sector” pada tipe residen menjadi 2 dan seterusnya.

Akhiran .1 ditambahkan karena di variable pelanggan sudah ada nama kolom yang sama. Sebagai contoh kolom “Jenis.Kelamin” yang terdapat pada pelanggan_matrix sudah ada juga di variable pelanggan. Jadi ketika digabungkan, R akan menambahkan akhiran .1 untuk kolom “Jenis.Kelamin” yang terdapat di pelanggan_matrix.

Terlihat data telah diringkaskan dengan teks kategori beserta pasangan numeriknya.

2. CLUSTERING DAN K-MEANS ALGORITHM

```
#Fungsi K-Means
field_yang_digunakan = c("Jenis.Kelamin.1", "Umur", "Profesi.1", "Tipe.Residen.1", "NilaiBelanjaSetahun")

#Bagian K-Means
set.seed(1)
#fungsi kmeans untuk membentuk 5 cluster dengan 25 skenario random dan simpan ke dalam variable segmentasi
segmentasi <- kmeans(x=pelanggan[field_yang_digunakan][c("Jenis.Kelamin.1", "Umur", "Profesi.1", "Tipe.Residen.1", "NilaiBelanjaSetahun")], nclusters=5)
#tampilkan hasil k-means
segmentasi

## K-means clustering with 5 clusters of sizes 12, 10, 5, 14, 9
##
## Cluster means:
##   Jenis.Kelamin.1   Umur Profesi.1 Tipe.Residen.1 NilaiBelanjaSetahun
## 1          1.75 31.58333  3.916667      1.250000      7.330958
## 2          1.70 52.50000  3.800000      1.300000      6.018321
```

```
## 3          1.40 61.80000  4.200000      1.400000      8.696132
## 4          2.00 20.07143  3.571429      1.357143      5.901089
## 5          2.00 42.33333  4.000000      1.555556      8.804791
##
## Clustering vector:
## [1] 3 4 2 2 5 4 3 2 1 1 2 2 3 3 4 1 1 3 1 4 5 2 1 5 1 2 1 5 2 5 4 5 4 4 5 1 4 5
## [39] 4 4 4 1 1 4 4 4 2 5 1 2
##
## Within cluster sum of squares by cluster:
## [1] 174.85164 108.49735  58.21123 316.73367 171.67372
## (between_SS / total_SS =  92.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"       "
```

```
#Analisa Hasil Clustering Vector - Penggabungan hasil cluster
segmentasi$cluster
```

```
## [1] 3 4 2 2 5 4 3 2 1 1 2 2 3 3 4 1 1 3 1 4 5 2 1 5 1 2 1 5 2 5 4 5 4 4 5 1 4 5
## [39] 4 4 4 1 1 4 4 4 2 5 1 2
```

```
pelanggan$cluster <- segmentasi$cluster
str(pelanggan)
```

```
## 'data.frame': 50 obs. of 11 variables:
## $ Customer_ID : chr "CUST-001" "CUST-002" "CUST-003" "CUST-004" ...
## $ Nama.Pelanggan : chr "Budi Anggara" "Shirley Ratuwati" "Agus Cahyono" "Antonius Winarta" ...
## $ Jenis.Kelamin : chr "Pria" "Wanita" "Pria" "Pria" ...
## $ Umur : int 58 14 48 53 41 24 64 52 29 33 ...
## $ Profesi : chr "Wiraswasta" "Pelajar" "Professional" "Professional" ...
## $ Tipe.Residen : chr "Sector" "Cluster" "Cluster" "Cluster" ...
## $ NilaiBelanjaSetahun: num 9.5 2.72 5.29 5.2 10.62 ...
## $ Jenis.Kelamin.1 : int 1 2 1 1 2 2 1 1 2 1 ...
## $ Profesi.1 : int 5 3 4 4 5 4 5 4 4 4 ...
## $ Tipe.Residen.1 : int 2 1 1 1 1 1 2 1 2 1 ...
## $ cluster : int 3 4 2 2 5 4 3 2 1 1 ...
```

```
#Analisa Hasil Cluster Size
#List data pada Cluster
which(pelanggan$cluster == 1)
```

```
## [1] 9 10 16 17 19 23 25 27 36 42 43 49
```

```
which(pelanggan$cluster == 2)
```

```
## [1] 3 4 8 11 12 22 26 29 47 50
```

```
which(pelanggan$cluster == 3)
```

```
## [1] 1 7 13 14 18
```

```
which(pelanggan$cluster == 4)
```

```
## [1] 2 6 15 20 31 33 34 37 39 40 41 44 45 46
```

```
which(pelanggan$cluster == 5)
```

```
## [1] 5 21 24 28 30 32 35 38 48
```

```
#Jumlah data pada Cluster
```

```
length(which(pelanggan$cluster == 1))
```

```
## [1] 12
```

```
length(which(pelanggan$cluster == 2))
```

```
## [1] 10
```

```
length(which(pelanggan$cluster == 3))
```

```
## [1] 5
```

```
length(which(pelanggan$cluster == 4))
```

```
## [1] 14
```

```
length(which(pelanggan$cluster == 5))
```

```
## [1] 9
```

```
#Melihat Data pada Cluster ke-N
```

```
#Melihat data cluster ke 3-5
```

```
pelanggan[which(pelanggan$cluster == 3),]
```

```
##      Customer_ID Nama.Pelanggan Jenis.Kelamin Umur      Profesi Tipe.Residen
## 1      CUST-001   Budi Anggara      Pria    58      Wiraswasta      Sector
## 7      CUST-007  Cahyono, Agus      Pria    64      Wiraswasta      Sector
## 13     CUST-013   Cahaya Putri      Wanita   64      Wiraswasta      Cluster
## 14     CUST-014  Mario Setiawan      Pria    60      Wiraswasta      Cluster
## 18     CUST-018   Nelly Halim      Wanita   63 Ibu Rumah Tangga      Cluster
##      NilaiBelanjaSetahun Jenis.Kelamin.1 Profesi.1 Tipe.Residen.1 cluster
## 1              9.497927              1          5              2          3
## 7              9.837260              1          5              2          3
## 13             9.333168              2          5              1          3
## 14             9.471615              1          5              1          3
## 18             5.340690              2          1              1          3
```

```
pelanggan[which(pelanggan$cluster == 4),]
```

```
##      Customer_ID  Nama.Pelanggan Jenis.Kelamin Umur      Profesi
## 2      CUST-002 Shirley Ratuwati      Wanita  14      Pelajar
## 6      CUST-006 Rosalina Kurnia      Wanita  24      Professional
## 15     CUST-015 Shirley Ratuwati      Wanita  20      Wiraswasta
## 20     CUST-020 Irene Novianto      Wanita  16      Pelajar
## 31     CUST-031 Eviana Handry      Wanita  19      Mahasiswa
## 33     CUST-033 Cecilia Kusnadi      Wanita  19      Mahasiswa
## 34     CUST-034 Deasy Arisandi      Wanita  21      Wiraswasta
## 37     CUST-037 Felicia Tandiono      Wanita  25      Professional
## 39     CUST-039 Gina Hidayat      Wanita  20      Professional
## 40     CUST-040 Irene Darmawan      Wanita  14      Pelajar
## 41     CUST-041 Shinta Aritonang      Wanita  24      Ibu Rumah Tangga
## 44     CUST-044      Anna      Wanita  18      Wiraswasta
## 45     CUST-045 Rismawati Juni      Wanita  22      Professional
## 46     CUST-046 Elfira Surya      Wanita  25      Wiraswasta
##      Tipe.Residen NilaiBelanjaSetahun Jenis.Kelamin.1 Profesi.1 Tipe.Residen.1
## 2      Cluster      2.722700      2      3      1
## 6      Cluster      5.215541      2      4      1
## 15     Cluster     10.365668      2      5      1
## 20     Sector      2.896845      2      3      2
## 31     Cluster      3.042773      2      2      1
## 33     Cluster      3.047926      2      2      1
## 34     Sector      9.759822      2      5      2
## 37     Sector      5.972787      2      4      2
## 39     Cluster      5.257775      2      4      1
## 40     Sector      2.861855      2      3      2
## 41     Cluster      6.820976      2      1      1
## 44     Cluster      9.339737      2      5      1
## 45     Cluster      5.211041      2      4      1
## 46     Sector     10.099807      2      5      2
##      cluster
## 2      4
## 6      4
## 15     4
## 20     4
## 31     4
## 33     4
## 34     4
## 37     4
## 39     4
## 40     4
## 41     4
## 44     4
## 45     4
## 46     4
```

```
pelanggan[which(pelanggan$cluster == 5),]
```

```
##      Customer_ID      Nama.Pelanggan Jenis.Kelamin Umur      Profesi
## 5      CUST-005 Ibu Sri Wahyuni, IR      Wanita  41      Wiraswasta
```

```

## 21    CUST-021    Lestari Fabianto    Wanita    38    Wiraswasta
## 24    CUST-024    Putri Ginting    Wanita    39    Wiraswasta
## 28    CUST-028    Adeline Huang    Wanita    40    Ibu Rumah Tangga
## 30    CUST-030    Rosita Saragih    Wanita    46    Ibu Rumah Tangga
## 32    CUST-032    Chintya Winarni    Wanita    47    Wiraswasta
## 35    CUST-035    Ida Ayu    Wanita    39    Professional
## 38    CUST-038    Agatha Salim    Wanita    46    Wiraswasta
## 48    CUST-048    Maria Hutagalung    Wanita    45    Wiraswasta
##      Tipe.Residen NilaiBelanjaSetahun Jenis.Kelamin.1 Profesi.1 Tipe.Residen.1
## 5      Cluster      10.615206      2      5      1
## 21      Cluster      9.222070      2      5      1
## 24      Cluster     10.259572      2      5      1
## 28      Cluster      6.631680      2      1      1
## 30      Sector      5.020976      2      1      2
## 32      Sector     10.663179      2      5      2
## 35      Sector      5.962575      2      4      2
## 38      Sector     10.477127      2      5      2
## 48      Sector     10.390732      2      5      2
##      cluster
## 5      5
## 21      5
## 24      5
## 28      5
## 30      5
## 32      5
## 35      5
## 38      5
## 48      5

```

```

#Analisa Hasil Cluster Means
segmentasi$centers

```

```

##      Jenis.Kelamin.1      Umur Profesi.1 Tipe.Residen.1 NilaiBelanjaSetahun
## 1      1.75 31.58333 3.916667      1.250000      7.330958
## 2      1.70 52.50000 3.800000      1.300000      6.018321
## 3      1.40 61.80000 4.200000      1.400000      8.696132
## 4      2.00 20.07143 3.571429      1.357143      5.901089
## 5      2.00 42.33333 4.000000      1.555556      8.804791

```

```

#Analisa Hasil Sum of Squares
#Membandingkan dengan 2 cluster kmeans, masing-masing 2 dan 5
set.seed(1)
kmeans(x=pelanggan[field_yang_digunakan], centers=2, nstart=25)

```

```

## K-means clustering with 2 clusters of sizes 23, 27
##
## Cluster means:
##      Jenis.Kelamin.1      Umur Profesi.1 Tipe.Residen.1 NilaiBelanjaSetahun
## 1      1.739130 51.17391 3.913043      1.434783      7.551518
## 2      1.888889 25.85185 3.777778      1.296296      6.659586
##
## Clustering vector:
## [1] 1 2 1 1 1 2 1 1 2 2 1 1 1 1 2 2 2 1 2 2 2 1 2 1 2 1 2 1 1 1 2 1 2 2 1 2 2 1

```

```
## [39] 2 2 2 2 2 2 2 2 1 1 2 1
##
## Within cluster sum of squares by cluster:
## [1] 1492.481 1524.081
## (between_SS / total_SS = 72.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"       "
```

```
set.seed(1)
kmeans(x=pelanggan[field_yang_digunakan], centers=5, nstart=25)
```

```
## K-means clustering with 5 clusters of sizes 12, 10, 5, 14, 9
##
## Cluster means:
##   Jenis.Kelamin.1   Umur Profesi.1 Tipe.Residen.1 NilaiBelanjaSetahun
## 1                1.75 31.58333  3.916667        1.250000        7.330958
## 2                1.70 52.50000  3.800000        1.300000        6.018321
## 3                1.40 61.80000  4.200000        1.400000        8.696132
## 4                2.00 20.07143  3.571429        1.357143        5.901089
## 5                2.00 42.33333  4.000000        1.555556        8.804791
##
## Clustering vector:
## [1] 3 4 2 2 5 4 3 2 1 1 2 2 3 3 4 1 1 3 1 4 5 2 1 5 1 2 1 5 2 5 4 5 4 4 5 1 4 5
## [39] 4 4 4 1 1 4 4 4 2 5 1 2
##
## Within cluster sum of squares by cluster:
## [1] 174.85164 108.49735  58.21123 316.73367 171.67372
## (between_SS / total_SS = 92.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"       "
```

```
#Available Components
segmentasi$withinss
```

```
## [1] 174.85164 108.49735  58.21123 316.73367 171.67372
```

```
segmentasi$cluster
```

```
## [1] 3 4 2 2 5 4 3 2 1 1 2 2 3 3 4 1 1 3 1 4 5 2 1 5 1 2 1 5 2 5 4 5 4 4 5 1 4 5
## [39] 4 4 4 1 1 4 4 4 2 5 1 2
```

```
segmentasi$tot.withinss
```

```
## [1] 829.9676
```

Hasil ini dapat dibagi dalam lima bagian, dengan penjelasan sebagai berikut:

1. Ukuran / Jumlah titik data pada tiap cluster
2. Nilai rata-rata (centroid) dari tiap cluster
3. Pembagian cluster dari tiap elemen data berdasarkan posisinya
4. Jumlah jarak kuadrat dari tiap titik ke centroidnya
5. Komponen informasi lain yang terkandung di dalam objek ini

K-means telah membagi dataset pelanggan dengan 5 cluster dengan jumlah total 50 data, yang juga merupakan jumlah data total pelanggan., dimana:

1. Cluster ke-1 memiliki 12 data
2. Cluster ke-2 memiliki 10 data
3. Cluster ke-3 memiliki 5 data
4. Cluster ke-4 memiliki 14 data
5. Cluster ke-5 memiliki 9 data

Cluster means adalah hasil nilai rata-rata atau titik sentral (centroid) dari seluruh titik tiap cluster. Pada tabel diatas:

- Kolom pertama yang berisi angka 1 sampai dengan 5 adalah mewakili nomor cluster.
- Kolom Kelamin.1 menunjukkan nilai rata-rata dari data jenis kelamin yang telah dikonversi menjadi numerik, dengan angka 1 mewakili Pria dan angka 2 mewakili wanita.

Penjabaran Cluster

1. Pada cluster 1 kolom Jenis.Kelamin.1 terlihat bahwa nilainya adalah 1.75, artinya data bersifat campuran namun cenderung ke wanita (2). Nah, untuk cluster ke-3 berupa angka 1.40 artinya data bersifat campuran namun cenderung ke Pria (1). Lalu pada cluster ke 4 dan 5 bernilai 2.00 yang berarti cluster hanya berisi data profil berjenis kelamin wanita.
2. Kolom Umur adalah representasi dari dataset awal tanpa mengalami konversi. Terlihat untuk cluster ke-1 umur rata-rata adalah 31 tahun, umur 52 tahun untuk cluster ke-2, dan seterusnya.
3. Kolom Profesi.1 menunjukkan nilai rata-rata data Profesi untuk tiap cluster yang telah dikonversi menjadi numerik, yaitu angka 1 s/d 5. Angka 1, 2, 3, 4, dan 5 di kolom ini masing-masingnya secara berurutan mewakili Ibu Rumah Tangga, Mahasiswa, Pelajar, Professional, dan Wiraswasta. Terlihat untuk seluruh cluster bahwa nilai profesi berada dalam rentang 3.5 s/d 4.2 ($3.5 < \text{profesi} \leq 4.2$). Hal ini menunjukkan bahwa profesi cenderung ke ke Professional, dan secara tegas cluster kelima memiliki profesi berupa Professional.
4. Kolom Tipe.Residen.1 menunjukkan representasi data Tipe.Residen yang telah dikonversi menjadi numerik dengan angka 1 mewakili Cluster dan 2 mewakili Sector. Ini juga didapatkan dari hasil konversi data menjadi numerik pada praktek sebelumnya. Untuk seluruh cluster, terlihat data cukup tersebar antara Sector dan Cluster terutama untuk cluster ke-5 dimana nilai kolom ini di angka 1.555.
5. Terakhir, kolom NilaiBelanjaSetahun cukup signifikan pembagiannya untuk tiap cluster. Cluster ke-3 dan ke-5 memiliki nilai belanja lebih tinggi dibandingkan ketiga cluster lainnya.

3. DATA VISUALIZATION

```
#install package  
install.packages("ggplot2")
```

```
## Installing package into 'C:/Users/sofyan/AppData/Local/R/win-library/4.3'  
## (as 'lib' is unspecified)
```

```
## package 'ggplot2' successfully unpacked and MD5 sums checked  
##  
## The downloaded binary packages are in  
## C:\Users\sofyan\AppData\Local\Temp\RtmpKeHfHl\downloaded_packages
```

```
#import library  
library(ggplot2)
```

```
#Simulasi Jumlah Cluster dan SS
```

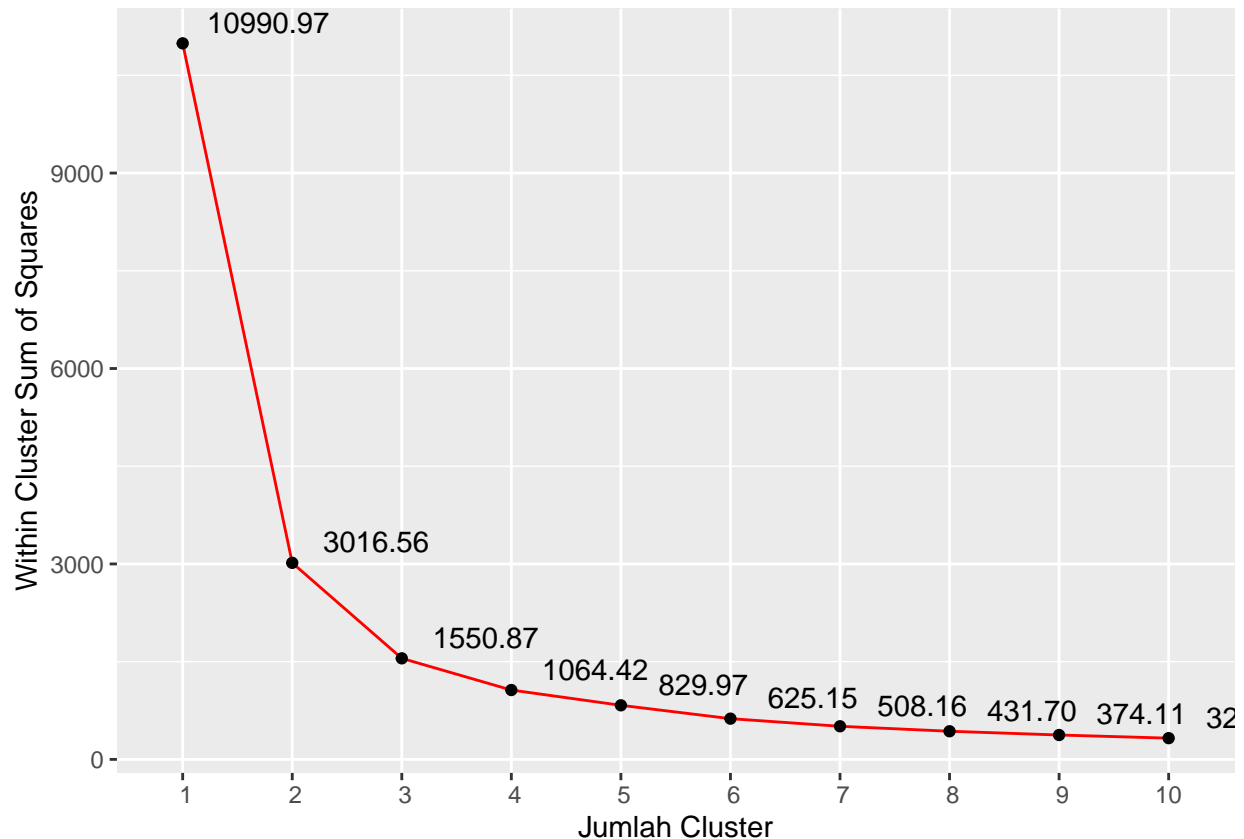
```
set.seed(1)  
sse <- sapply(1:10,  
             function(param_k){  
               kmeans(pelanggan[field_yang_digunakan], param_k, nstart=25)$tot.withinss  
             })  
sse
```

```
## [1] 10990.9740 3016.5612 1550.8725 1064.4187 829.9676 625.1462  
## [7] 508.1568 431.6977 374.1095 325.7982
```

```
#Grafik Elbow Effect
```

```
jumlah_cluster_max <- 10  
ssdata = data.frame(cluster=c(1:jumlah_cluster_max),sse)  
ggplot(ssdata, aes(x=cluster,y=sse)) +  
  geom_line(color="red") + geom_point() +  
  ylab("Within Cluster Sum of Squares") + xlab("Jumlah Cluster") +  
  geom_text(aes(label=format(round(sse, 2), nsmall = 2)),hjust=-0.2, vjust=-0.5) +  
  scale_x_discrete(limits=c(1:jumlah_cluster_max))
```

```
## Warning: Continuous limits supplied to discrete scale.  
## i Did you mean 'limits = factor(...)' or 'scale*_continuous()'?
```

Metrik elbow method yang digunakan sebagai basis justifikasi adalah Sum of Squares (SS), atau lebih tepatnya komponen **tot.withinss** (total penjumlahan dari tiap SS dari withinss) dari **objek kmeans**.

Metrik ini akan dicari progressive nilai tot.withinss untuk tiap kombinasi jumlah cluster, dan disimpan dalam bentuk vector di R.

Untuk keperluan ini, function sapply digunakan. Function sapply akan digunakan untuk memanggil function kmeans untuk suatu range jumlah cluster. Range ini akan kita gunakan 1 sampai dengan 10.

Terlihat jika jumlah cluster optimal yang bisa digunakan adalah 5, dan ini menjadi keputusan untuk menentukan jumlah segmentasi pelanggan.

4. PEMAKETAN MODEL K-MEANS

```
#Menamakan Segmen
Segmen.Pelanggan <- data.frame(cluster=c(1,2,3,4,5), Nama.Segmen=c("Silver Youth Gals", "Diamond Senior

#Menggabungkan Referensi
Identitas.Cluster <- list(Profesi=Profesi, Jenis.Kelamin=Jenis.Kelamin, Tipe.Residen=Tipe.Residen, Segmen=Segmen

#Menyimpan Objek dalam Bentuk File
saveRDS(Identitas.Cluster,"cluster.rds")
```

Penamaan Cluster

- Cluster 1 : Silver Youth Gals: alasannya adalah karena umurnya rata-rata adalah 20, wanita semua, profesinya bercampur antar pelajar dan professional serta pembelanjaan sekitar 6 juta.
- Cluster 2 : Diamond Senior Member: alasannya adalah karena umurnya rata-rata adalah 61 tahun dan pembelanjaan di atas 8 juta.
- Cluster 3 : Gold Young Professional: alasannya adalah karena umurnya rata-rata adalah 31 tahun, professional dan pembelanjaan cukup besar.
- Cluster 4 : Diamond Profesional: alasannya adalah karena umurnya rata-rata adalah 42 tahun, pembelanjaan paling tinggi dan semuanya professional.
- Cluster 5 : Silver Mid Professional: alasannya adalah karena umurnya rata-rata adalah 52 tahun dan pembelanjaan sekitar 6 juta.

Untuk menyimpan objek yang telah dibuat ke dalam file digunakan function saveRDS. File ini kemudian dapat dibuka kembali sebagai objek ke depannya. Objek yang sudah digabungkan sebelumnya sudah memiliki semua aset yang diperlukan untuk mengalokasikan data baru ke segmen yang sesuai.

5. INPUT DATA BARU

```
#Data Baru
```

```
databaru <- data.frame(Customer_ID="CUST-100", Nama.Pelanggan="Rudi Wilamar",Umur=20,Jenis.Kelamin="Wanita",
databaru
```

```
## Customer_ID Nama.Pelanggan Umur Jenis.Kelamin Profesi Tipe.Residen
## 1 CUST-100 Rudi Wilamar 20 Wanita Pelajar Cluster
## NilaiBelanjaSetahun
## 1 3.5
```

```
#Memuat Objek Clustering dari File
```

```
Identitas.Cluster <- readRDS(file = "cluster.rds")
print(Identitas.Cluster)
```

```
## $Profesi
## Profesi Profesi.1
## 1 Wiraswasta 5
## 2 Pelajar 3
## 3 Professional 4
## 17 Ibu Rumah Tangga 1
## 31 Mahasiswa 2
##
## $Jenis.Kelamin
## Jenis.Kelamin Jenis.Kelamin.1
## 1 Pria 1
## 2 Wanita 2
##
## $Tipe.Residen
## Tipe.Residen Tipe.Residen.1
## 1 Sector 2
## 2 Cluster 1
```

```
##
## $Segmentasi
## K-means clustering with 5 clusters of sizes 12, 10, 5, 14, 9
##
## Cluster means:
##   Jenis.Kelamin.1   Umur Profesi.1 Tipe.Residen.1 NilaiBelanjaSetahun
## 1             1.75 31.58333  3.916667      1.250000      7.330958
## 2             1.70 52.50000  3.800000      1.300000      6.018321
## 3             1.40 61.80000  4.200000      1.400000      8.696132
## 4             2.00 20.07143  3.571429      1.357143      5.901089
## 5             2.00 42.33333  4.000000      1.555556      8.804791
##
## Clustering vector:
## [1] 3 4 2 2 5 4 3 2 1 1 2 2 3 3 4 1 1 3 1 4 5 2 1 5 1 2 1 5 2 5 4 5 4 4 5 1 4 5
## [39] 4 4 4 1 1 4 4 4 2 5 1 2
##
## Within cluster sum of squares by cluster:
## [1] 174.85164 108.49735  58.21123 316.73367 171.67372
## (between_SS / total_SS =  92.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
##
## $Segmen.Pelanggan
##   cluster      Nama.Segmen
## 1         1   Silver Youth Gals
## 2         2   Diamond Senior Member
## 3         3 Gold Young Professional
## 4         4   Diamond Professional
## 5         5 Silver Mid Professional
##
## $field_yang_digunakan
## [1] "Jenis.Kelamin.1"      "Umur"              "Profesi.1"
## [4] "Tipe.Residen.1"      "NilaiBelanjaSetahun"
```

#Merge dengan Data Referensi

#Masukkan perintah untuk penggabungan data

```
databaru <- merge(databaru, Identitas.Cluster$Profesi)
databaru <- merge(databaru, Identitas.Cluster$Jenis.Kelamin)
databaru <- merge(databaru, Identitas.Cluster$Tipe.Residen)
databaru
```

```
##   Tipe.Residen Jenis.Kelamin Profesi Customer_ID Nama.Pelanggan Umur
## 1      Cluster      Wanita Pelajar    CUST-100   Rudi Wilamar   20
##   NilaiBelanjaSetahun Profesi.1 Jenis.Kelamin.1 Tipe.Residen.1
## 1             3.5           3           2           1
```

#Menentukan Cluster

#menentukan data baru di cluster mana

```
which.min(sapply( 1:5, function( x ) sum( ( databaru[Identitas.Cluster$field_yang_digunakan] - Identitas
```

```
## [1] 4
```

```
Identitas.Cluster$Segmen.Pelanggan[which.min(sapply( 1:5, function( x ) sum( ( databaru[Identitas.Clust
```

```
##      cluster      Nama.Segmen  
## 4          4 Diamond Professional
```

Dengan menggunakan model dan paket yang telah dibuat sebelumnya, maka dapat dilakukan uji data baru dengan memuat object clustering yang sudah ada.

Data baru yang dimasukkan : Customer__ID="CUST-100", Nama.Pelanggan="Rudi Wilamar", Umur=20, Jenis.Kelamin="W"

Data baru kemudian di merge dengan referensi yang ada kemudian dapat ditentukan cluster mana yang sesuai dengan data baru yang dimasukkan tersebut dengan tahapan berikut:

- Mencari jarak kuadrat minimum atau terdekat
- Dari kolom numerik data baru tersebut
- Ke centroid kolom terkait
- Dari seluruh cluster yang ada

Dimana:

- min: Merupakan function untuk mencari nilai minimum
- 1:5 : Adalah range nomor cluster dari 1 sampai dengan 5 (atau lebih sesuai dengan ukuran cluster)
- sapply: Digunakan untuk melakukan iterasi berdasarkan range (dalam kasus ini 1 s/d 5)
- function(x): Digunakan untuk proses dengan x diisi 1 s/d 5 per proses
- (data[kolom] – objekkmeans\$centers[x,]) ^2: Adalah jarak kuadrat data. Ingat centers adalah komponen dari objek kmeans.
- sum: Digunakan untuk menjumlahkan jarak kuadrat

Yang kemudian didapatkan hasil **Cluster 4 Diamond Professional**.