

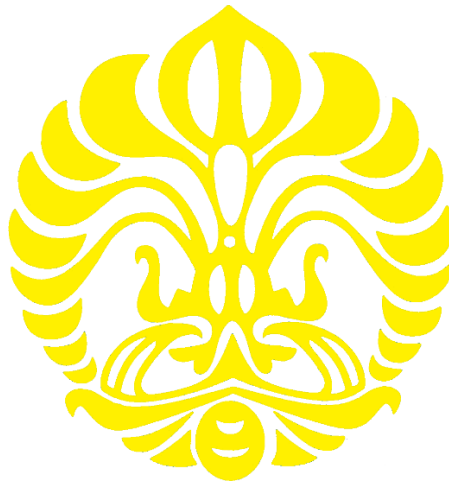
UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



PROJECT AKHIR ANALISIS DATA



Disusun oleh:

Michael Mario Bramanthyo Adhi 1906299534

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS INDONESIA
DEPOK
2022**

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Tujuan:

Untuk mempelajari dan mengetahui segmentasi klien kartu kredit dan mempelajari pola konsumsi klien. Segmentasi klien adalah strategi pemasaran yang mengelompokkan target pasar berdasarkan karakteristik yang sama untuk dikelola secara efektif dan tepat agar mencapai tujuan bisnis yang telah ditetapkan.

Data Set

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data yang berisikan informasi klien kartu kredit di suatu bank. Dengan bantuan *software* Rstudio akan ditampilkan 6 data pertama dari segmentasi pelanggan.

CUST_ID	BALANCE	FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FREQUENCY	PURCHASES_INSTALLMENTS_FREQUENCY	CASH_ADVANCE_FREQUENCY	CASH_ADVANCE_TRX	PURCHASES_TRX	CREDIT_LIMIT	PAYMENTS	MINIMUM_PAYMENTS	PRC_FULL_PAYMENT	TENURE
C10001	40,900749	0,818182	95,4	0	95,4	0	0,166667	0	0,083333	0	0	2	1000	201,802084	136,509787	0	12
C10002	3202,467416	0,909091	0	0	0	6442,945483	0	0	0	0,25	4	0	7000	4108,032597	1072,340217	0,222222	12
C10003	2495,148862	1	773,17	773,17	0	0	1	1	0	0	0	12	7500	622,066742	627,284787	0	12
C10004	1666,670542	0,636364	1499	1499	0	205,788017	0,083333	0,083333	0	0,083333	1	1	7500	0	0	0	12
C10005	817,714335	1	16	16	0	0	0,083333	0,083333	0	0	0	1	1200	678,334763	244,791237	0	12
C10006	1809,828751	1	1333,28	0	1333,28	0	0,666667	0	0,583333	0	0	8	1800	1400,05777	2407,246035	0	12

Berikut merupakan variabel-variabel yang terdapat pada data segmentasi pelanggan

```
> dim(df)
[1] 8950 17
> names(df)
[1] "BALANCE" "BALANCE_FREQUENCY"
[3] "PURCHASES" "ONEOFF_PURCHASES"
[5] "INSTALLMENTS_PURCHASES" "CASH_ADVANCE"
[7] "PURCHASES_FREQUENCY" "ONEOFF_PURCHASES_FREQUENCY"
[9] "PURCHASES_INSTALLMENTS_FREQUENCY" "CASH_ADVANCE_FREQUENCY"
[11] "CASH_ADVANCE_TRX" "PURCHASES_TRX"
[13] "CREDIT_LIMIT" "PAYMENTS"
[15] "MINIMUM_PAYMENTS" "PRC_FULL_PAYMENT"
[17] "TENURE"
```

Berdasarkan *output* tersebut diketahui bahwa data segmentasi pelanggan berdimensi [8950, 17], yang mana diinterpretasikan sebagai 8950 observasi dan 17 variabel observasi.

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Penjelasan Variabel

CUSTID : id customer (Categorical)

BALANCE : sisa saldo dalam rekening

BALANCE_FREQUENCY : intensitas perubahan saldo, antara 0 dan 1 (1 = sering, 0 = jarang)

PURCHASES : nilai total pembelian yang dilakukan

ONEOFF_PURCHASES : nilai pembelian maksimum dalam satu transaksi one-go

INSTALLMENTS_PURCHASES : nilai pembelian dengan cicilan yang pernah dilakukan

CASH_ADVANCE : DP/uang muka yang dibayarkan

PURCHASES_FREQUENCY : intensitas pembelian, antara 0 dan 1 (1 = sering, 0 = jarang)

ONEOFF_PURCHASES_FREQUENCY : intensitas pembelian one-go (1 = sering, 0 = jarang)

PURCHASES_INSTALLMENTS_FREQUENCY : intensitas pembelian dengan cicilan (1 = sering, 0 = jarang)

CASH_ADVANCE_FREQUENCY : seberapa sering DP/uang muka dibayarkan

CASH_ADVANCE_TRX : banyaknya transaksi dengan DP/uang muka

PURCHASES_TRX : banyaknya transaksi pembelian

CREDIT_LIMIT : limit kartu kredit

PAYMENTS : banyaknya pembayaran yang dilakukan

MINIMUM_PAYMENTS : nilai minimal pembayaran


PRCFULL_PAYMENT : persentase pembayaran penuh yang dilakukan

TENURE : masa tenur kartu kredit

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Berikut informasi terkait variabel yang terdapat pada data segmentasi pelanggan

```
> str(df)
tibble [8,950 x 17] (S3: tbl_df/tbl/data.frame)
 $ BALANCE                : num [1:8950] 40.9 3202.5 2495.1 1666.7 817.7
 ...
 $ BALANCE_FREQUENCY      : num [1:8950] 0.818 0.909 1 0.636 1 ...
 $ PURCHASES              : num [1:8950] 95.4 0 773.2 1499 16 ...
 $ ONEOFF_PURCHASES       : num [1:8950] 0 0 773 1499 16 ...
 $ INSTALLMENTS_PURCHASES : num [1:8950] 95.4 0 0 0 0 ...
 $ CASH_ADVANCE           : num [1:8950] 0 6443 0 206 0 ...
 $ PURCHASES_FREQUENCY    : num [1:8950] 0.1667 0 1 0.0833 0.0833 ...
 $ ONEOFF_PURCHASES_FREQUENCY : num [1:8950] 0 0 1 0.0833 0.0833 ...
 $ PURCHASES_INSTALLMENTS_FREQUENCY : num [1:8950] 0.0833 0 0 0 0 ...
 $ CASH_ADVANCE_FREQUENCY : num [1:8950] 0 0.25 0 0.0833 0 ...
 $ CASH_ADVANCE_TRX       : num [1:8950] 0 4 0 1 0 0 0 0 0 ...
 $ PURCHASES_TRX         : num [1:8950] 2 0 12 1 1 8 64 12 5 3 ...
 $ CREDIT_LIMIT           : num [1:8950] 1000 7000 7500 7500 1200 1800 13500
 2300 7000 11000 ...
 $ PAYMENTS              : num [1:8950] 202 4103 622 0 678 ...
 $ MINIMUM_PAYMENTS      : num [1:8950] 140 1072 627 NA 245 ...
 $ PRC_FULL_PAYMENT      : num [1:8950] 0 0.222 0 0 0 ...
 $ TENURE                : num [1:8950] 12 12 12 12 12 12 12 12 12 ...
```

Berdasarkan *output* tersebut diketahui bahwa data segmentasi pelanggan terdiri atas variabel numerik dan variabel kategorik, serta beberapa variabel terdapat *missing value*.

Handling Missing Value

Dengan menggunakan *software* Rstudio diperoleh:

```
> colSums(is.na(df))
      BALANCE      BALANCE_FREQUENCY
      0          0
      PURCHASES      ONEOFF_PURCHASES
      0          0
INSTALLMENTS_PURCHASES      CASH_ADVANCE
      0          0
      PURCHASES_FREQUENCY      ONEOFF_PURCHASES_FREQUENCY
      0          0
PURCHASES_INSTALLMENTS_FREQUENCY      CASH_ADVANCE_FREQUENCY
      0          0
      CASH_ADVANCE_TRX      PURCHASES_TRX
      0          0
      CREDIT_LIMIT      PAYMENTS
      1          0
      MINIMUM_PAYMENTS      PRC_FULL_PAYMENT
      313          0
      TENURE
      0
```

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Berdasarkan *output* tersebut diperoleh bahwa variabel **CREDIT_LIMIT** dan **MINIMUM_PAYMENTS** terdapat *missing value*, maka perlu dilakukan penanganan pada variabel tersebut.

Berikut data hasil *handling missing value*.

```
> dim(df1)
[1] 8636 17
```

Setelah dilakukan *handling missing value* diperoleh bahwa data segmentasi pelanggan berdimensi [8636, 17], yang mana diinterpretasikan sebagai 8636 observasi dan 17 variabel observasi.

Diperoleh *summary* dari data setelah dilakukan *handling missing value* adalah

```
> summary(df1)
      BALANCE      BALANCE_FREQUENCY      PURCHASES      ONEOFF_PURCHASES
Min.   : 0.0    Min.   :0.0000    Min.   : 0.00    Min.   : 0.00
1st Qu.: 148.1  1st Qu.:0.9091    1st Qu.: 43.37  1st Qu.: 0.00
Median : 916.9  Median :1.0000    Median : 375.40  Median : 44.99
Mean   : 1601.2 Mean   :0.8950    Mean   : 1025.43  Mean   : 604.90
3rd Qu.: 2105.2 3rd Qu.:1.0000    3rd Qu.: 1145.98  3rd Qu.: 599.10
Max.   :19043.1 Max.   :1.0000    Max.   :49039.57  Max.   :40761.25
INSTALLMENTS_PURCHASES      CASH_ADVANCE      PURCHASES_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
Min.   : 0.00    Min.   : 0.0    Min.   :0.000000    Min.   :0.000000
1st Qu.: 0.00    1st Qu.: 0.0    1st Qu.:0.08333    1st Qu.:0.000000
Median : 94.78    Median : 0.0    Median :0.50000    Median :0.08333
Mean   : 420.84    Mean   : 994.2    Mean   :0.49600    Mean   :0.20591
3rd Qu.: 484.15    3rd Qu.: 1132.4  3rd Qu.:0.91667    3rd Qu.:0.33333
Max.   :22500.00    Max.   :47137.2  Max.   :1.00000    Max.   :1.00000
PURCHASES_INSTALLMENTS_FREQUENCY      CASH_ADVANCE_FREQUENCY      CASH_ADVANCE TRX
Min.   :0.00000    Min.   :0.0000    Min.   : 0.000
1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.: 0.000
Median :0.1667    Median :0.0000    Median : 0.000
Mean   :0.3688    Mean   :0.1376    Mean   : 3.314
3rd Qu.:0.7500    3rd Qu.:0.2500    3rd Qu.: 4.000
Max.   :1.0000    Max.   :1.5000    Max.   :123.000
PURCHASES_TRX      CREDIT_LIMIT      PAYMENTS      MINIMUM_PAYMENTS      PRC_FULL_PAYMENT
Min.   : 0.00    Min.   : 50    Min.   : 0.05    Min.   : 0.02    Min.   :0.0000
1st Qu.: 1.00    1st Qu.: 1600  1st Qu.: 418.56  1st Qu.: 169.16  1st Qu.:0.0000
Median : 7.00    Median : 3000  Median : 896.68  Median : 312.45  Median :0.0000
Mean   : 15.03    Mean   : 4522  Mean   : 1784.48  Mean   : 864.30  Mean   :0.1593
3rd Qu.: 18.00    3rd Qu.: 6500  3rd Qu.: 1951.14  3rd Qu.: 825.50  3rd Qu.:0.1667
Max.   :358.00    Max.   :30000  Max.   :50721.48  Max.   :76406.21  Max.   :1.0000
TENURE
Min.   : 6.00
1st Qu.:12.00
Median :12.00
Mean   :11.53
3rd Qu.:12.00
Max.   :12.00
```

UAS Analisis Data

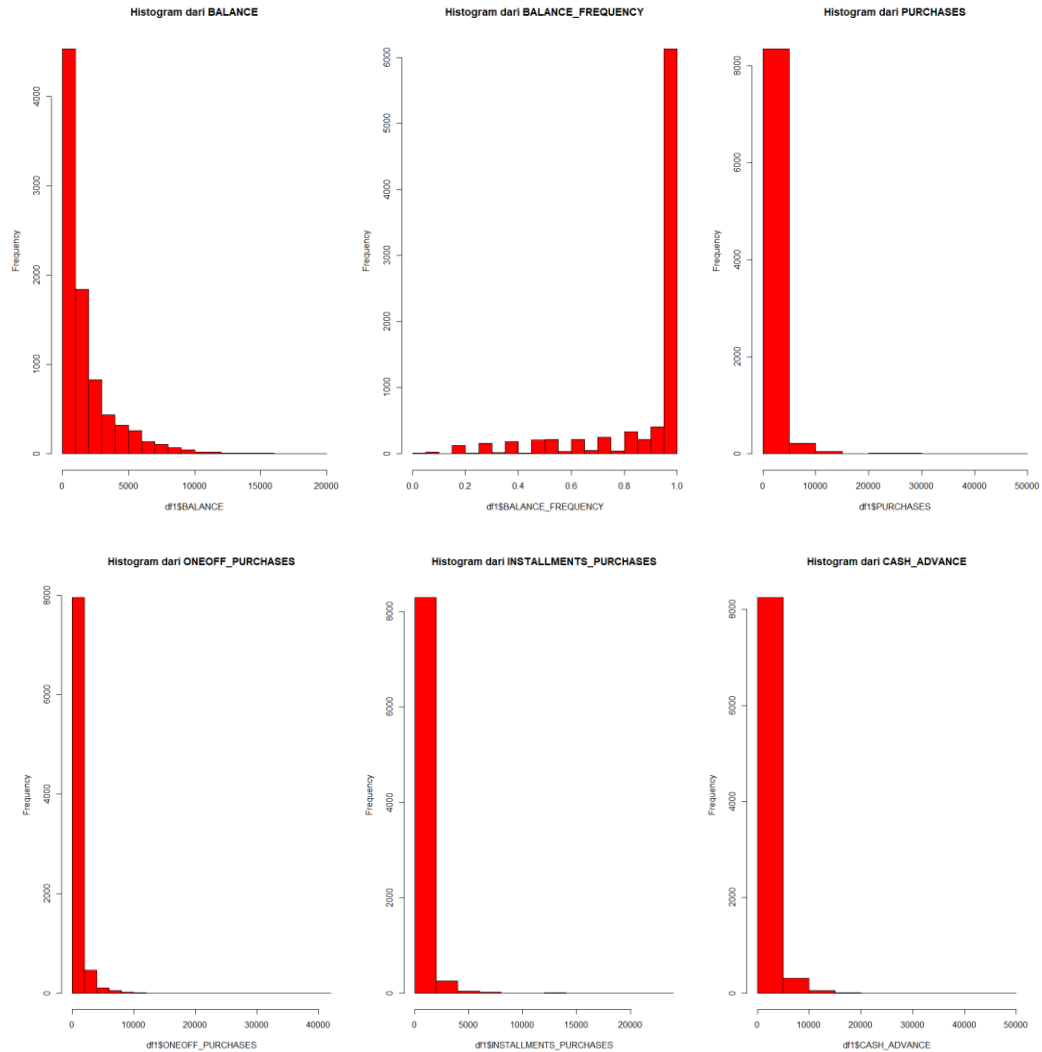
Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Visualisasi Data

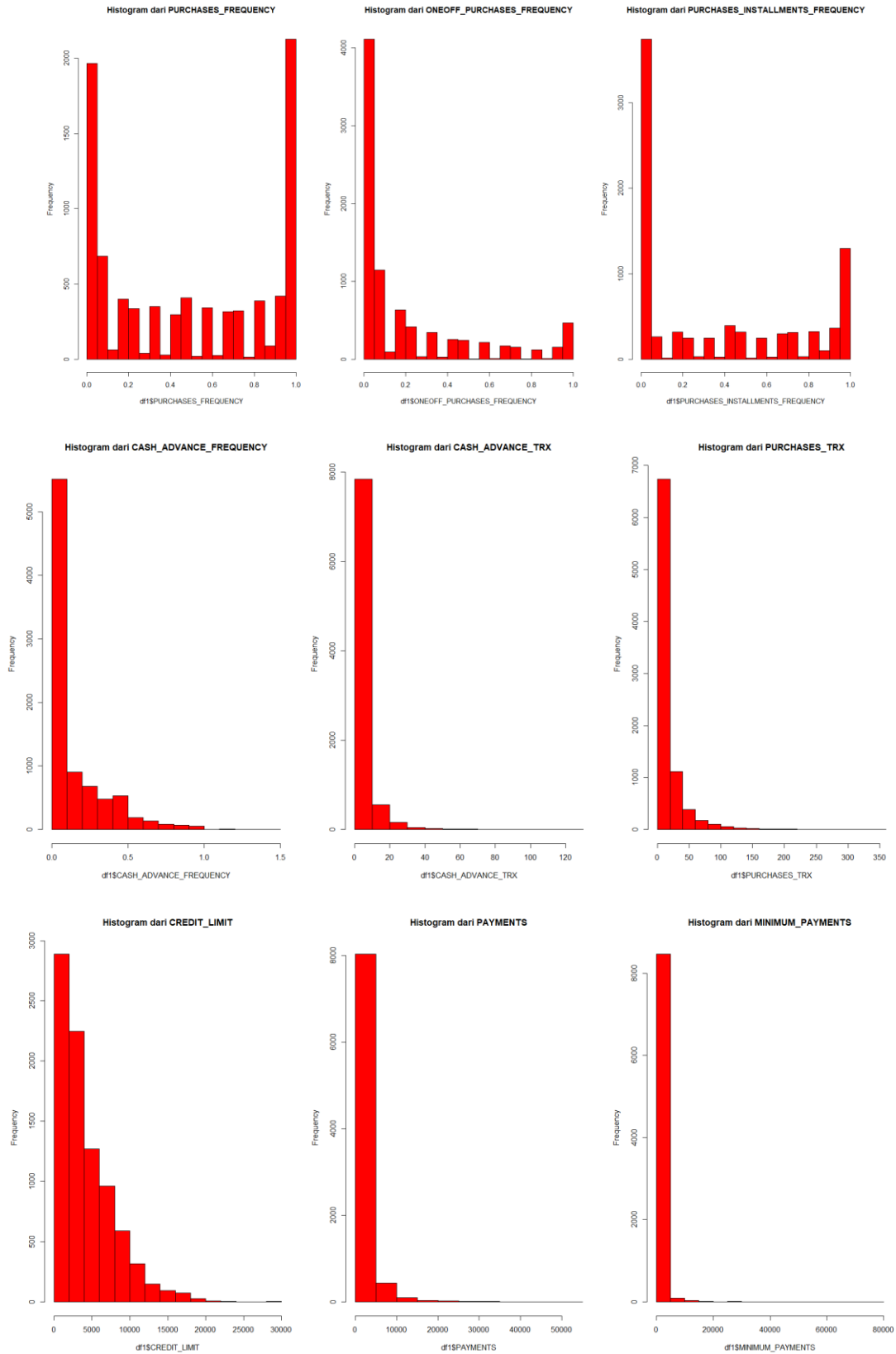
Histogram



UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

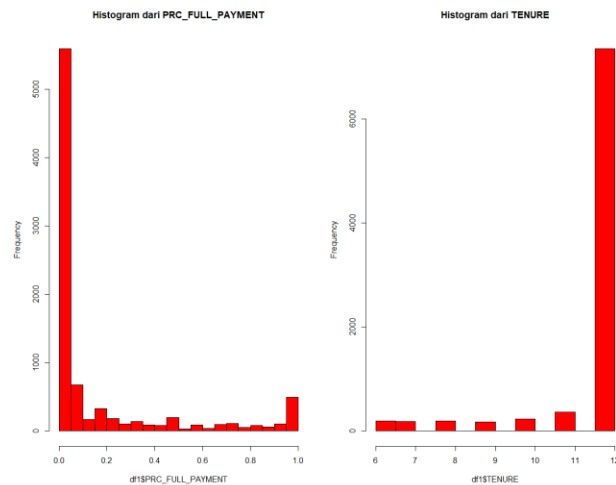
NPM : 1906299534



UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Berdasarkan histogram di atas dapat diperoleh informasi bahwa sisa saldo dalam rekening terbesar berada dalam interval 0-1000, intensitas perubahan saldo sering terjadi, nilai total pembelian yang dilakukan terbesar berada dalam interval 0-5000, nilai pembelian maksimum dalam satu transaksi one-go terbesar berada dalam interval 0-2000, nilai pembelian dengan cicilan yang pernah dilakukan terbesar berada dalam interval 0-1000, DP/uang muka yang dibayarkan terbesar berada dalam interval 0-5000, intensitas pembelian sering terjadi, intensitas pembelian one-go jarang terjadi, intensitas pembelian dengan cicilan jarang terjadi, DP/uang muka dibayarkan jarang terjadi, banyaknya transaksi dengan DP/uang muka terbesar berada dalam interval 0-10, banyaknya transaksi pembelian terbesar berada dalam interval 0-20, limit kartu kredit terbesar berada dalam interval 0-2000, banyaknya pembayaran yang dilakukan terbesar berada dalam interval 0-5000, nilai minimal pembayaran terbesar berada dalam interval 0-5000, persentase pembayaran penuh yang dilakukan terbesar berada dalam interval 0-5%, dan masa tenur kartu kredit terlama berada dalam interval 12 bulan

UAS Analisis Data

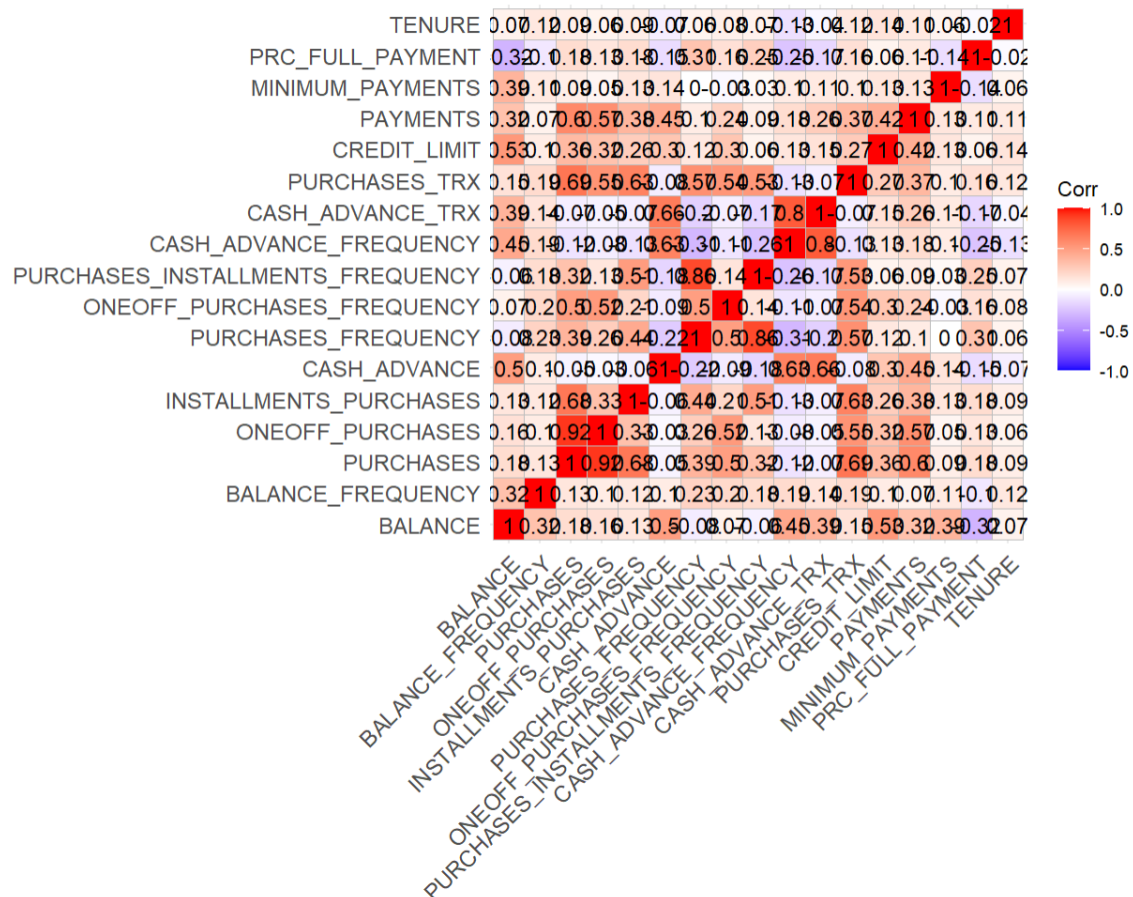
Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Plot Korelasi

Dengan menggunakan *software* Rstudio diperoleh plot korelasi antar variabel numerik



Nilai koefisien korelasi (r) dapat bervariasi dari -1 sampai +1. Nilai r yang mendekati -1 atau +1 menunjukkan hubungan yang kuat antara dua variabel tersebut dan nilai r yang mendekati 0 mengindikasikan lemahnya hubungan antara dua variabel tersebut.

Berdasarkan plot di atas diperoleh bahwa PURCHASES dengan ONEOFF_PURCHASES, CASH_ADVANCE_FREQUENCY dengan CASH_ADVANCE_TRX, dan PURCHASE_INSTALLMENTS_FREQUENCY dengan PURCHASES_FREQUENCY memiliki korelasi yang positif dan menunjukkan hubungan linier yang kuat karena nilai korelasinya yang mendekati +1. Hal ini menunjukkan bahwa semakin besar nilai total pembelian yang dilakukan jika semakin besar juga nilai pembelian maksimum dalam satu transaksi one-go. Semakin sering DP/uang muka dibayarkan jika semakin besar banyaknya

UAS Analisis Data

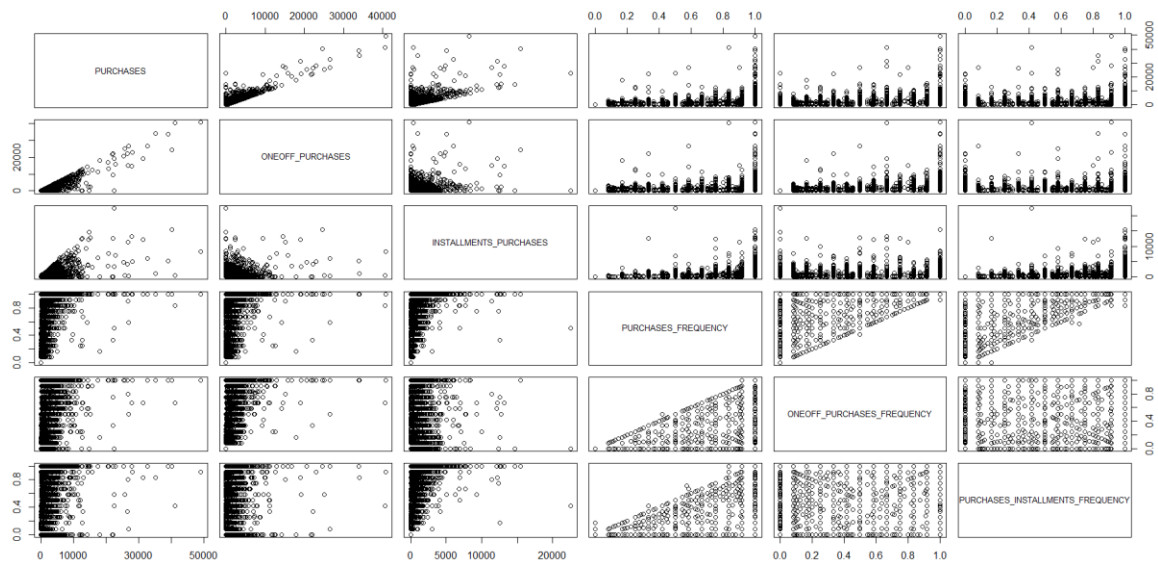
Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



transaksi dengan DP/uang muka. Selain itu, semakin sering intensitas pembelian dengan cicilan jika semakin sering intensitas pembelian.

Scatter Plot Matriks Korelasi



Berdasarkan *scatter plot* di atas dapat diperoleh informasi bahwa:

1. Terdapat hubungan linier positif antara PURCHASES dengan ONEOFF_PURCHASES yang menunjukkan bahwa semakin besar nilai total pembelian yang dilakukan jika semakin besar juga nilai pembelian maksimum dalam satu transaksi one-go
2. Terdapat hubungan linier positif antara PURCHASES dengan INSTALLMENTS_PURCHASES yang menunjukkan bahwa semakin besar nilai total pembelian yang dilakukan jika semakin besar juga nilai pembelian dengan cicilan yang pernah dilakukan
3. Terdapat hubungan linier positif antara PURCHASES_FREQUENCY dengan ONEOFF_PURCHASES_FREQUENCY yang menunjukkan bahwa semakin sering intensitas pembelian jika semakin sering juga intensitas pembelian one-go
4. Terdapat hubungan linier positif antara PURCHASES_FREQUENCY dengan PURCHASES_INSTALLMENTS_FREQUENCY yang menunjukkan bahwa semakin

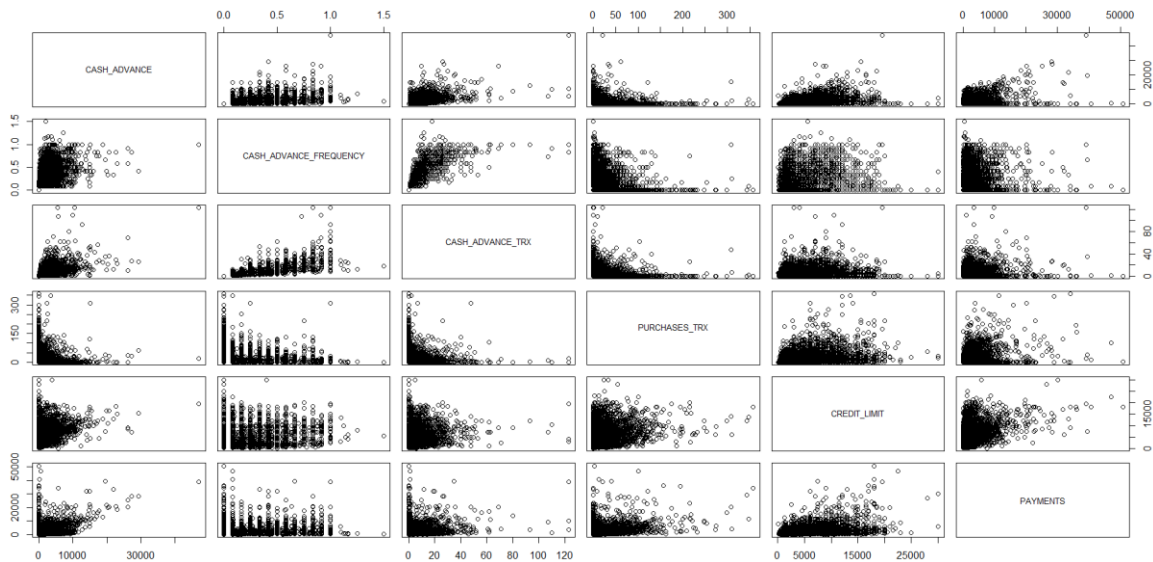
UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



sering intensitas pembelian jika semakin sering juga intensitas pembelian dengan cicilan



Berdasarkan *scatter plot* di atas dapat diperoleh informasi bahwa:

1. Terdapat hubungan linier positif antara CASH_ADVANCE dengan CASH_ADVANCE_TRX yang menunjukkan bahwa semakin besar DP/uang muka yang dibayarkan jika semakin besar juga banyaknya transaksi dengan DP/uang muka
2. Terdapat hubungan linier positif antara CASH_ADVANCE_FREQUENCY dengan CASH_ADVANCE_TRX yang menunjukkan bahwa semakin sering DP/uang muka dibayarkan jika semakin besar juga banyaknya transaksi dengan DP/uang muka
3. Terdapat hubungan linier positif antara CREDIT_LIMIT dengan PAYMENTS yang menunjukkan bahwa semakin besar limit kartu kredit jika semakin besar juga banyaknya pembayaran yang dilakukan

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Untuk mendapatkan informasi mengenai segmentasi klien akan dilakukan analisis data dengan menggunakan metode *principal component analysis* (PCA), analisis faktor (FA), dan analisis cluster (CA). *Principal Component Analysis* (PCA) digunakan untuk mereduksi variabel-variabel sehingga mengurangi kompleksitas hubungan timbal-balik antara sejumlah besar variabel yang diamati ke sejumlah relatif kecil dari kombinasi linearnya, yang disebut sebagai komponen utama. *Factor Analysis* (FA) digunakan untuk mengidentifikasi hubungan antar variabel dengan menggunakan teknik *Exploratory Factor Analysis*, serta akan dilakukan analisis apakah data dapat direduksi ke dalam beberapa faktor dengan menggunakan *Confirmatory Factor Analysis*. Sedangkan, *Cluster Analysis* (CA) digunakan untuk mengklasifikasi objek atau kasus (responden) ke dalam kelompok yang relatif homogen.

Analisis Data

1. *Principal Component Analysis* (PCA)

Tujuan:

Untuk mereduksi variabel-variabel sehingga mengurangi kompleksitas hubungan timbal-balik antara sejumlah besar variabel yang diamati ke sejumlah relatif kecil dari kombinasi linearnya, yang disebut sebagai komponen utama.

Langkah Kerja:

1. Standarisasi data
2. Mencari matriks kovariansi dari data
3. Mencari vektor eigen dan nilai eigen dari matriks kovariansi atau matriks korelasi, atau melakukan *Singular Value Decomposition*
4. Mengurutkan nilai eigen dalam urutan menurun
5. Menghitung proporsi dari variansi serta kumulatif dari proporsi
6. Menentukan jumlah komponen utama berdasarkan kumulatif dari proporsi
7. Membentuk persamaan yang dibentuk dari kombinasi linier antara vektor eigen dengan data yang telah distandarisasi

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Standardisasi Data

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_i}{s_i}, \quad i = 1, 2, \dots, 18$$

Standarisasi data dilakukan untuk mengubah nilai asli data menjadi bentuk distribusi normal. Standarisasi data dapat menggunakan function R `scale()` pada *dataframe*:

```
> ### Standarisasi Data ###
> scaled.data <- scale(df1)
> head(scaled.data)
```

	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES
[1,]	-0.7445817	-0.37002536	-0.4291590	-0.35913949	-0.3548054
[2,]	0.7641079	0.06767501	-0.4731808	-0.35913949	-0.4588125
[3,]	0.4265777	0.50537539	-0.1164058	0.09990369	-0.4588125
[4,]	-0.3738888	0.50537539	-0.4657977	-0.34964003	-0.4588125
[5,]	0.0995451	0.50537539	0.1420539	-0.35913949	0.9947574
[6,]	-0.4647726	0.50537539	2.7989266	3.44220252	0.2916739

	CASH_ADVANCE	PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FREQUENCY
[1,]	-0.4686284	-0.8207213	-0.6862398
[2,]	2.5684078	-1.2360673	-0.6862398
[3,]	-0.4686284	1.2560039	2.6464980
[4,]	-0.4686284	-1.0283956	-0.4085128
[5,]	-0.4686284	0.4253143	-0.6862398
[6,]	-0.4686284	1.2560039	2.6464980

	PURCHASES_INSTALLMENTS	FREQUENCY	CASH_ADVANCE_FREQUENCY	CASH_ADVANCE_TRX
[1,]	-0.7171374	-0.6819130	-0.47940912	
[2,]	-0.9264679	0.5569899	0.09925221	
[3,]	-0.9264679	-0.6819130	-0.47940912	
[4,]	-0.9264679	-0.6819130	-0.47940912	
[5,]	0.5388507	-0.6819130	-0.47940912	
[6,]	1.5855083	-0.6819130	-0.47940912	

	PURCHASES_TRX	CREDIT_LIMIT	PAYMENTS	MINIMUM_PAYMENTS	PRC_FULL_PAYMENT	TENURE
[1,]	-0.5175930	-0.9625197	-0.5439104	-0.30548994	-0.5376958	0.3551601
[2,]	-0.5970196	0.6771649	0.7968061	0.08768365	0.2123677	0.3551601
[3,]	-0.1204598	0.8138052	-0.3994801	-0.09990033	-0.5376958	0.3551601
[4,]	-0.5573063	-0.9078636	-0.3801428	-0.26111544	-0.5376958	0.3551601
[5,]	-0.2793130	-0.7438951	-0.1321118	0.65032579	-0.5376958	0.3551601
[6,]	1.9446329	2.4534898	1.5704929	-0.28080945	2.8375934	0.3551601

Matriks Kovariansi

$$S = \frac{1}{n-1} \mathbf{X}' \left(\mathbf{I} - \frac{1}{n} \mathbf{J} \right) \mathbf{X}$$

Untuk mencari matriks kovarians dapat digunakan function R `cov()` pada *dataframe* yang sudah distandarisasi:

```
> #Covariance Matrix
> covariance.matrix <- cov(scaled.data)
> covariance.matrix
```

	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES
BALANCE	1.00000000	0.31014045	0.17608309	0.15998462
BALANCE_FREQUENCY	0.31014045	1.00000000	0.12263475	0.09525421
PURCHASES	0.17608309	0.12263475	1.00000000	0.91678022
ONEOFF_PURCHASES	0.15998462	0.09525421	0.91678022	1.00000000
INSTALLMENTS_PURCHASES	0.12210872	0.11473904	0.67925890	0.32964963
CASH_ADVANCE	0.49558562	0.08903585	-0.05376022	-0.03324436
PURCHASES_FREQUENCY	-0.08845944	0.22815802	0.39299969	0.26546048

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



ONEOFF_PURCHASES_FREQUENCY	0.06383163	0.18746742	0.49738407	0.52451406
PURCHASES_INSTALLMENTS_FREQUENCY	-0.06958199	0.18415944	0.31602493	0.12837981
CASH_ADVANCE_FREQUENCY	0.44530716	0.18113216	-0.12486256	-0.08641311
CASH_ADVANCE_TRX	0.38238845	0.13326522	-0.07027723	-0.04870521
PURCHASES_TRX	0.14788721	0.18309471	0.68873161	0.54531267
CREDIT_LIMIT	0.53551818	0.08768229	0.35842540	0.32061345
PAYMENTS	0.32283000	0.03916910	0.60678236	0.57085035
MINIMUM_PAYMENTS	0.39866892	0.13251925	0.09384200	0.04874124
PRC_FULL_PAYMENT	-0.33359378	-0.15696080	0.17644722	0.12988997
TENURE	0.06698709	0.10471441	0.08454519	0.06339986
INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY		
BALANCE	0.12210872	0.49558562	-0.088459439	
BALANCE_FREQUENCY	0.11473904	0.08903585	0.228158017	
PURCHASES	0.67925890	-0.05376022	0.392999691	
ONEOFF_PURCHASES	0.32964963	-0.03324436	0.265460483	
INSTALLMENTS_PURCHASES	1.00000000	-0.06609907	0.441193440	
CASH_ADVANCE	-0.06609907	1.00000000	-0.218566480	
PURCHASES_FREQUENCY	0.44119344	-0.21856648	1.000000000	
ONEOFF_PURCHASES_FREQUENCY	0.21182881	-0.09011083	0.502123268	
PURCHASES_INSTALLMENTS_FREQUENCY	0.51113030	-0.17887343	0.862337912	
CASH_ADVANCE_FREQUENCY	-0.13645481	0.62903025	-0.316770718	
CASH_ADVANCE_TRX	-0.07671580	0.65691147	-0.208749159	
PURCHASES_TRX	0.62608254	-0.07844868	0.567168259	
CREDIT_LIMIT	0.25805749	0.30416069	0.121371988	
PAYMENTS	0.38539118	0.45934244	0.100715175	
MINIMUM_PAYMENTS	0.13215579	0.14009387	0.002978715	
PRC_FULL_PAYMENT	0.17843959	-0.15856634	0.305393309	
TENURE	0.08334567	-0.07308979	0.052160820	
ONEOFF_PURCHASES_FREQUENCY	PURCHASES_INSTALLMENTS_FREQUENCY			
BALANCE	0.06383163	-0.06958199		
BALANCE_FREQUENCY	0.18746742	0.18415944		
PURCHASES	0.49738407	0.31602493		
ONEOFF_PURCHASES	0.52451406	0.12837981		
INSTALLMENTS_PURCHASES	0.21182881	0.51113030		
CASH_ADVANCE	-0.09011083	-0.17887343		
PURCHASES_FREQUENCY	0.50212327	0.86233791		
ONEOFF_PURCHASES_FREQUENCY	1.00000000	0.14391127		
PURCHASES_INSTALLMENTS_FREQUENCY	0.14391127	1.00000000		
CASH_ADVANCE_FREQUENCY	-0.11822491	-0.26892396		
CASH_ADVANCE_TRX	-0.07338372	-0.17281090		
PURCHASES_TRX	0.54436419	0.52919156		
CREDIT_LIMIT	0.29543234	0.06342088		
PAYMENTS	0.24434588	0.08342412		
MINIMUM_PAYMENTS	-0.03035650	0.03003625		
PRC_FULL_PAYMENT	0.15407317	0.24954659		
TENURE	0.08043720	0.06577935		
CASH_ADVANCE_FREQUENCY	CASH_ADVANCE_TRX	PURCHASES_TRX		
BALANCE	0.44530716	0.38238845	0.14788721	
BALANCE_FREQUENCY	0.18113216	0.13326522	0.18309471	
PURCHASES	-0.12486256	-0.07027723	0.68873161	
ONEOFF_PURCHASES	-0.08641311	-0.04870521	0.54531267	
INSTALLMENTS_PURCHASES	-0.13645481	-0.07671580	0.62608254	
CASH_ADVANCE	0.62903025	0.65691147	-0.07844868	
PURCHASES_FREQUENCY	-0.31677072	-0.20874916	0.56716826	
ONEOFF_PURCHASES_FREQUENCY	-0.11822491	-0.07338372	0.54436419	
PURCHASES_INSTALLMENTS_FREQUENCY	-0.26892396	-0.17281090	0.52919156	
CASH_ADVANCE_FREQUENCY	1.00000000	0.79959285	-0.13691326	
CASH_ADVANCE_TRX	0.79959285	1.00000000	-0.06987920	
PURCHASES_TRX	-0.13691326	-0.06987920	1.00000000	
CREDIT_LIMIT	0.13284059	0.15026109	0.27371510	
PAYMENTS	0.18233951	0.25592143	0.37147739	
MINIMUM_PAYMENTS	0.09884446	0.10972697	0.09631068	
PRC_FULL_PAYMENT	-0.26000117	-0.17657414	0.15678887	
TENURE	-0.14003762	-0.04535010	0.11888947	
CREDIT_LIMIT	PAYMENTS	MINIMUM_PAYMENTS	PRC_FULL_PAYMENT	
BALANCE	0.53551818	0.32283000	0.398668920	-0.33359378
BALANCE_FREQUENCY	0.08768229	0.03916910	0.132519246	-0.15696080
PURCHASES	0.35842540	0.60678236	0.093842000	0.17644722
ONEOFF_PURCHASES	0.32061345	0.57085035	0.048741241	0.12988997
INSTALLMENTS_PURCHASES	0.25805749	0.38539118	0.132155787	0.17843959
CASH_ADVANCE	0.30416069	0.45934244	0.140093869	-0.15856634
PURCHASES_FREQUENCY	0.12137199	0.10071517	0.002978715	0.30539331
ONEOFF_PURCHASES_FREQUENCY	0.29543234	0.24434588	-0.030356501	0.15407317
PURCHASES_INSTALLMENTS_FREQUENCY	0.06342088	0.08342412	0.030036248	0.24954659
CASH_ADVANCE_FREQUENCY	0.13284059	0.18233951	0.098844462	-0.26000117
CASH_ADVANCE_TRX	0.15026109	0.25592143	0.109726968	-0.17657414
PURCHASES_TRX	0.27371510	0.37147739	0.096310676	0.15678887
CREDIT_LIMIT	1.00000000	0.42695139	0.126670946	0.05257792
PAYMENTS	0.42695139	1.00000000	0.126629505	0.10466810

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



MINIMUM_PAYMENTS	0.12667095	0.12662950	1.000000000	-0.14040421
PRC_FULL_PAYMENT	0.05257792	0.10466810	-0.140404207	1.000000000
TENURE	0.13646382	0.10298580	0.059330304	-0.02422260
TENURE				
BALANCE	0.06698709			
BALANCE_FREQUENCY	0.10471441			
PURCHASES	0.08454519			
ONEOFF_PURCHASES	0.06339986			
INSTALLMENTS_PURCHASES	0.08334567			
CASH_ADVANCE	-0.07308979			
PURCHASES_FREQUENCY	0.05216082			
ONEOFF_PURCHASES_FREQUENCY	0.08043720			
PURCHASES_INSTALLMENTS_FREQUENCY	0.06577935			
CASH_ADVANCE_FREQUENCY	-0.14003762			
CASH_ADVANCE_TRX	-0.04535010			
PURCHASES_TRX	0.11888947			
CREDIT_LIMIT	0.13646382			
PAYMENTS	0.10298580			
MINIMUM_PAYMENTS	0.05933030			
PRC_FULL_PAYMENT	-0.02422260			
TENURE	1.00000000			

Eigen Value dan Eigen Vector

Untuk mencari nilai eigen dan vektor eigen dapat digunakan function `eigen()` pada matriks kovarians:

```
> ### Eigen Value dan Eigen Vector ###
> eigen <- eigen(covariance.matrix)
> eigen
eigen() decomposition
$values
[1] 4.629300e+00 3.463632e+00 1.516117e+00 1.287513e+00 1.067025e+00 9.711736e-01 8.357519e-01
[8] 7.158248e-01 6.258884e-01 5.238563e-01 4.018466e-01 3.016717e-01 2.425407e-01 2.003722e-01
[15] 1.713765e-01 4.609765e-02 1.173078e-05

$vectors
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
[1,] -0.09198590 -0.405978695 -0.174155217 0.25942307 -0.07570040 0.03576313 0.26336950
[2,] -0.10981218 -0.127738729 -0.458853339 0.15932011 0.45085906 -0.01465339 -0.09867483
[3,] -0.41215123 -0.049530298 0.242581867 0.06400168 0.01041000 0.19599250 -0.20135723
[4,] -0.34677536 -0.069929651 0.368572605 0.12314791 0.19702123 0.17300614 -0.11273384
[5,] -0.33705564 0.011481319 -0.103753039 -0.07502838 -0.33748918 0.14543134 -0.26897198
[6,] 0.03058765 -0.437246881 -0.001725939 -0.26556462 -0.09942509 -0.13251141 0.03854000
[7,] -0.32366488 0.186581704 -0.355749762 -0.22173800 0.08853801 -0.08569339 0.15790086
[8,] -0.29476135 0.014746578 0.104743080 0.05546399 0.52155960 -0.09682787 0.30574575
[9,] -0.27722626 0.173576912 -0.449940259 -0.26529256 -0.17540752 -0.04745999 -0.04322889
[10,] 0.09914541 -0.429996888 -0.087635459 -0.26659223 0.15993226 0.03154756 -0.13731595
[11,] 0.05696036 -0.416411836 -0.087051918 -0.33264408 0.08974284 -0.08978691 -0.19670125
[12,] -0.39106653 0.011946595 -0.079799025 -0.02410640 0.05252272 0.07813781 -0.10423056
[13,] -0.21005184 -0.243823086 0.095181896 0.12272574 -0.13200136 -0.31286494 0.54388385
[14,] -0.26372547 -0.264181758 0.287920713 -0.09751655 -0.18919193 -0.06565158 -0.16879476
[15,] -0.05932632 -0.170415769 -0.248706216 0.35220392 -0.41681138 0.34027647 0.20417945
[16,] -0.13056503 0.195708905 0.184195976 -0.41815027 -0.20109740 -0.28866061 0.28035918
[17,] -0.07791867 0.004565576 -0.065743191 0.42837395 -0.11778693 -0.74566146 -0.40066207

      [,8]      [,9]      [,10]      [,11]      [,12]      [,13]      [,14]
[1,] 0.199909116 -0.06197751 0.04507583 -0.15051794 0.47603934 0.53776360 0.142931091
[2,] -0.127517815 -0.67124591 -0.02664969 0.13922295 -0.06735340 -0.16874883 -0.023207170
[3,] 0.004961326 -0.10127322 0.05933502 -0.19641144 -0.07900829 0.10871129 -0.224677700
[4,] -0.123272594 -0.06905673 -0.16512046 -0.44648142 0.04939272 -0.01078444 -0.223011102
[5,] 0.238348234 -0.11249796 0.44371023 0.35610940 -0.27747873 0.27661467 -0.121272113
[6,] 0.004672424 0.01914138 -0.37380405 0.35285930 0.17402761 0.01057757 -0.597409014
[7,] -0.026416556 0.19145556 -0.25789128 -0.12579403 -0.16148484 0.19351323 0.010277748
[8,] -0.199748785 0.36226652 0.08905699 0.37039216 -0.16552752 0.24779547 0.043774920
[9,] 0.129174475 0.08214839 -0.25554276 -0.29615331 0.01747642 -0.04083117 0.043562583
[10,] -0.077263491 0.08684470 0.29006015 -0.21298490 -0.03754873 0.04262180 0.337622654
[11,] -0.179753605 0.21481677 0.20775042 -0.20853964 -0.20389321 -0.09393053 -0.118324846
[12,] 0.045473232 0.25517812 0.22976565 0.20288582 0.59358733 -0.52975825 0.080454848
[13,] 0.367373919 -0.09414198 0.16063017 -0.15191194 -0.32059510 -0.40190308 -0.030390689
[14,] -0.048019266 -0.13562050 -0.45992657 0.26006200 -0.11797783 -0.04245051 0.604103939
[15,] -0.613456344 0.14800036 -0.01656729 0.02237652 -0.15951141 -0.14112581 -0.023690712
[16,] -0.481560271 -0.39284213 0.26595119 -0.04971503 0.24636254 0.11293777 0.008694622
```

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



```
[17,] -0.169092838  0.14356536  0.04119176 -0.06555575  0.03064761  0.07727909 -0.048262432
      [,15]      [,16]      [,17]
[1,]  0.21899594  0.005697928 -4.234793e-05
[2,]  0.04364865  0.009378625 -2.196411e-05
[3,] -0.06287646 -0.001484780  7.488622e-01
[4,] -0.06812328  0.005222760 -5.821123e-01
[5,] -0.02347596 -0.013983240 -3.167815e-01
[6,] -0.23754880  0.008379976 -2.182227e-05
[7,] -0.02388578 -0.678750052  1.595440e-04
[8,] -0.01549951  0.341692602 -1.420775e-04
[9,] -0.06734487  0.632813034 -1.687351e-04
[10,] -0.64739670 -0.041773789  1.562213e-05
[11,]  0.64854493  0.014367929 -1.122564e-05
[12,]  0.03815196 -0.134871073  8.297720e-05
[13,] -0.05389870 -0.018635564  8.652891e-06
[14,]  0.13785392 -0.011141115  1.564578e-05
[15,] -0.07201141  0.014566076 -2.867162e-05
[16,]  0.01067467  0.020994292 -2.046526e-05
[17,] -0.10400839 -0.020260848  3.486820e-06
```

Variance

Total Variance

$$\sum_{j=1}^{18} S_{jj} = 1 + 1 + \dots + 1 = 18$$

Dengan menggunakan *software* Rstudio, didapat *total variance* sebagai berikut:

```
> ### Total Variance ###
> sum(diag(covariance.matrix))
[1] 17
> sum(eigen$values)
[1] 17
```

Proportion of Variance

$$\text{Proportion of Variance} = \frac{\lambda_i}{\sum_{j=1}^{18} S_{jj}}, \quad i = 1, 2, \dots, 18$$

Dengan menggunakan *software* Rstudio, didapat *proportion of variance* sebagai berikut:

```
> ### Proportion of Variance ###
> prop.variance <- eigen$values/sum(eigen$values)
> prop.variance
[1] 2.723118e-01 2.037431e-01 8.918334e-02 7.573609e-02 6.276618e-02 5.712786e-02 4.916188e-02
[8] 4.210734e-02 3.681697e-02 3.081508e-02 2.363804e-02 1.774540e-02 1.426710e-02 1.178660e-02
[15] 1.008097e-02 2.711627e-03 6.900460e-07
```


UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Cummulative Proportion

$$\text{Cummulative Proportion} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{\sum_{j=1}^{18} S_{jj}}, \quad k = 1, 2, \dots, 18$$

Dengan menggunakan *software* Rstudio, didapat *cumulative proportion* sebagai berikut:

```
> ### Cumulative of Proportion ###  
> cumprop <- cumsum(prop.variance)  
> cumprop  
[1] 0.2723118 0.4760548 0.5652382 0.6409743 0.7037405 0.7608683 0.8100302 0.8521375 0.8889545  
[10] 0.9197696 0.9434076 0.9611530 0.9754201 0.9872067 0.9972877 0.9999993 1.0000000
```

Dengan menggunakan analisis komponen utama, kita dapat mereduksi dimensi dari data segmentasi klien. Setelah dilakukan analisis data, diperoleh bahwa dari 17 komponen utama yang ada kita akan mengambil hanya 5 komponen utama pertama. Berdasarkan *output* nilai eigen, dapat dilihat bahwa nilai eigen lebih besar dari 1 (> 1) adalah komponen utama pertama, yaitu PC1, PC2, PC3, PC4, dan PC5, hal ini tentunya akan mempercepat komputasi.

Kita juga dapat melihat proporsi varians dari nilai eigen. Apabila kita hanya menggunakan satu komponen utama, proporsi nilai eigennya hanya sebesar 27.231% atau dapat dikatakan masih belum menggambarkan variabilitas data. Namun, jika kita menggunakan 5 komponen utama, akumulasi proporsi variansnya adalah 0.6202. Dapat dikatakan keenam komponen utama yaitu PC1, PC2, PC3, PC4, PC5, dan PC6 dapat mewakili 70.37% dari total variabilitas data.

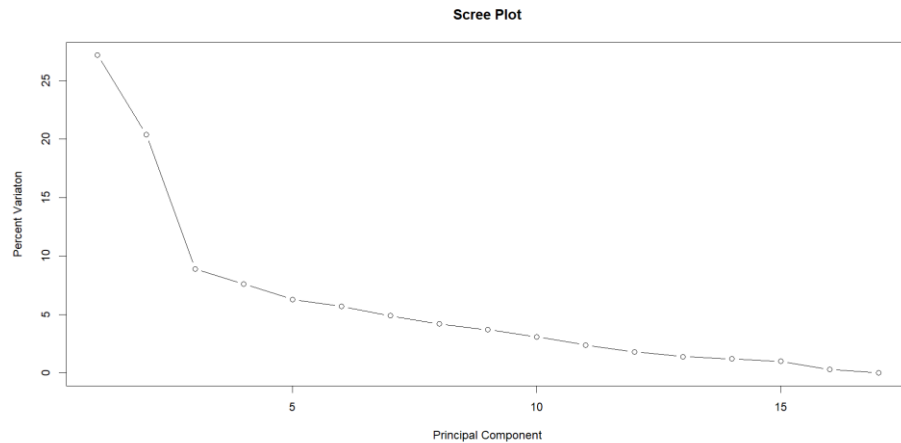
UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Scree Plot



Berdasarkan *Scree Plot* di atas, garis kurva mulai melandai setelah komponen utama ke-5. Maka berdasarkan *Scree Plot* kita dapat memilih 5 komponen utama yaitu PC1, PC2, PC3, PC4, dan PC5.

Untuk estimasi *factor loadings*, akan digunakan *principal component method* dengan rotasi varimax.

Factor Loadings (Facto Analysis Model)

$$\begin{aligned}\hat{\Lambda} &= \mathbf{C}_1 \mathbf{D}_1^{\frac{1}{2}} \\ \begin{bmatrix} \hat{\lambda}_{11} & \hat{\lambda}_{12} \\ \hat{\lambda}_{21} & \hat{\lambda}_{22} \\ \vdots & \vdots \\ \hat{\lambda}_{18\ 1} & \hat{\lambda}_{18\ 2} \end{bmatrix} &= \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ \vdots & \vdots \\ c_{18\ 1} & c_{18\ 2} \end{bmatrix} \begin{bmatrix} \sqrt{\theta_1} & 0 \\ 0 & \sqrt{\theta_2} \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{\theta_1} c_{11} & \sqrt{\theta_2} c_{12} \\ \sqrt{\theta_1} c_{21} & \sqrt{\theta_2} c_{22} \\ \vdots & \vdots \\ \sqrt{\theta_1} c_{18\ 1} & \sqrt{\theta_2} c_{18\ 2} \end{bmatrix}\end{aligned}$$

Varimax Rotation

$$\hat{\Lambda}^* = \hat{\Lambda} \mathbf{T}$$

Untuk mencari nilai *factor loadings* dapat digunakan *package* “psych”:

```
> ### Principal Component Analysis ###  
> fit <- principal(df1, x$ncomp, rotate="varimax")  
> print(fit$loadings, cutoff=.3)
```

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

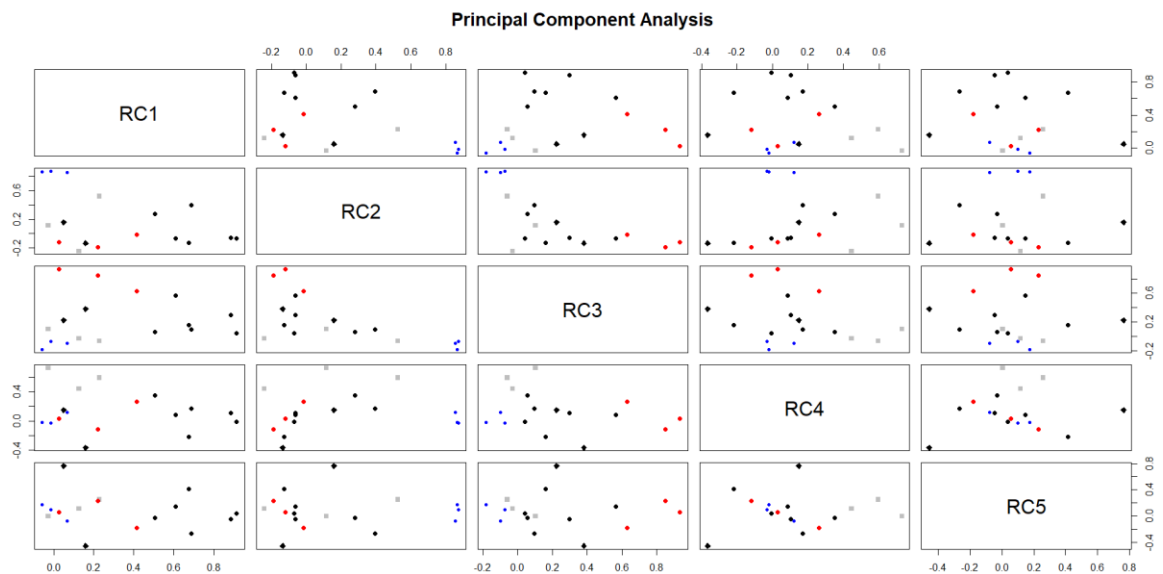
NPM : 1906299534



Loadings:	RC1	RC2	RC3	RC4	RC5
BALANCE		0.524		0.595	
BALANCE_FREQUENCY					0.763
PURCHASES	0.885				
ONEOFF_PURCHASES	0.912				
INSTALLMENTS_PURCHASES	0.415		0.631		
CASH_ADVANCE		0.857			
PURCHASES_FREQUENCY			0.848		
ONEOFF_PURCHASES_FREQUENCY	0.673				0.415
PURCHASES_INSTALLMENTS_FREQUENCY			0.934		
CASH_ADVANCE_FREQUENCY		0.866			
CASH_ADVANCE_TRX		0.873			
PURCHASES_TRX	0.608		0.566		
CREDIT_LIMIT	0.506			0.352	
PAYMENTS	0.688	0.396			
MINIMUM_PAYMENTS				0.729	
PRC_FULL_PAYMENT			0.381	-0.366	-0.456
TENURE				0.445	
SS loadings	3.491	2.948	2.698	1.556	1.271
Proportion Var	0.205	0.173	0.159	0.092	0.075
Cumulative Var	0.205	0.379	0.537	0.629	0.704

Diagram *Principal Component Analysis*

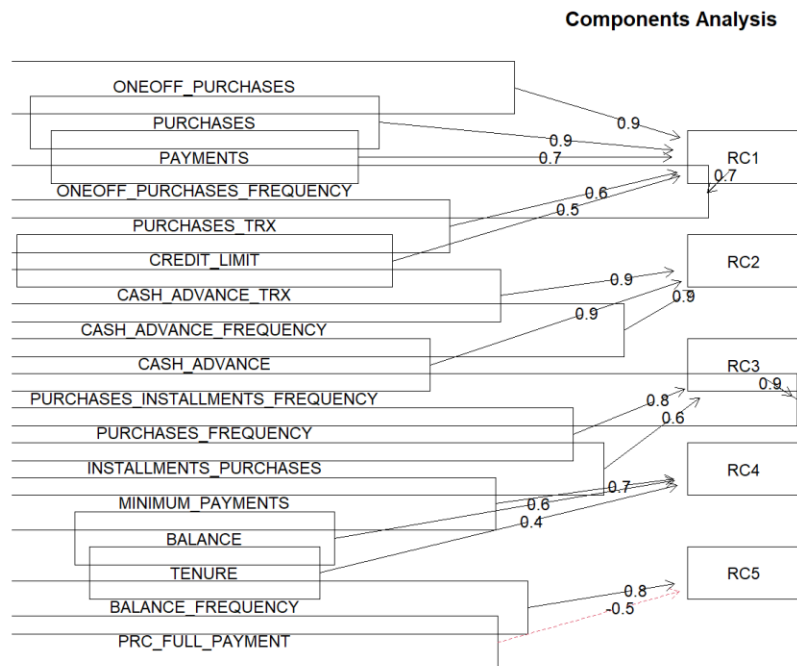
Dengan menggunakan *software* Rstudio diperoleh:



UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



∴ Jadi, berdasarkan *output factor loadings* dan diagram *component analysis* diperoleh 5 komponen utama diantaranya:

Komponen 1 : ONEOFF_PURCHASES, PURCHASES, PAYMENTS, ONEOFF_PURCHASES_FREQUENCY, PURCHASES_TRX, CREDIT_LIMIT

Komponen 2 : CASH_ADVANCE_TRX, CASH_ADVANCE_FREQUENCY, , CASH_ADVANCE

Komponen 3 : PURCHASES_INSTALLMENTS_FREQUENCY, PURCHASES_FREQUENCY, INSTALLMENTS_PURCHASES

Komponen 4 : MINIMUM_PAYMENTS, BALANCE, TENURE

Komponen 5 : BALANCE_FREQUENCY, PRC_FULL_PAYMENT

Model tersebut dapat menjelaskan 70.37% variansi dari data.

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



2. *Factor Analysis (FA)*

Tujuan:

Untuk mengidentifikasi hubungan antar variabel dengan menggunakan teknik *Exploratory Factor Analysis*, serta akan dilakukan analisis apakah data dapat direduksi ke dalam beberapa faktor dengan menggunakan *Confirmatory Factor Analysis*.

Untuk mereduksi data dengan mencari kombinasi linear dari variabel tersembunyi yang tidak diketahui atau faktor yang ingin dipelajari sekaligus mengidentifikasi apakah terdapat korelasi antar variabel. Hal tersebut dapat dimulai dengan mengasumsikan bahwa terdapat variabel yang membentuk variabel-variabel yang diukur dalam data.

Langkah Kerja

Exploratory Factor Analysis

1. Mencari matriks korelasi dari data
2. Menghitung nilai dan vektor eigen
3. Membuat *screeplot* untuk menentukan berapa faktor yang akan dibentuk
4. Estimasi loadings dan communalities menggunakan *principal component method*

Confirmatory Factor Analysis

1. Menentukan model
2. Menguji model baik atau tidak dengan *goodness of fit* Uji chi square, CFI, RMSEA, dan SRMR

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Standardisasi Data

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_i}{s_i}, \quad i = 1, 2, \dots, 18$$

Standarisasi data dilakukan untuk mengubah nilai asli data menjadi bentuk distribusi normal. Standarisasi data dapat menggunakan function R scale() pada *dataframe*:

```
> ### Standarisasi Data ###
> scaled.data <- scale(df1)
> head(scaled.data)
```

	BALANCE	BALANCE	FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES
[1,]	-0.7445817		-0.37002536	-0.4291590	-0.35913949	-0.3548054
[2,]	0.7641079		0.06767501	-0.4731808	-0.35913949	-0.4588125
[3,]	0.4265777		0.50537539	-0.1164058	0.09990369	-0.4588125
[4,]	-0.3738888		0.50537539	-0.4657977	-0.34964003	-0.4588125
[5,]	0.0995451		0.50537539	0.1420539	-0.35913949	0.9947574
[6,]	-0.4647726		0.50537539	2.7989266	3.44220252	0.2916739

	CASH_ADVANCE	PURCHASES	FREQUENCY	ONEOFF_PURCHASES	FREQUENCY
[1,]	-0.4686284		-0.8207213		-0.6862398
[2,]	2.5684078		-1.2360673		-0.6862398
[3,]	-0.4686284		1.2560039		2.6464980
[4,]	-0.4686284		-1.0283956		-0.4085128
[5,]	-0.4686284		0.4253143		-0.6862398
[6,]	-0.4686284		1.2560039		2.6464980

	PURCHASES_INSTALLMENTS	FREQUENCY	CASH_ADVANCE	FREQUENCY	CASH_ADVANCE	TRX
[1,]		-0.7171374		-0.6819130		-0.47940912
[2,]		-0.9264679		0.5569899		0.09925221
[3,]		-0.9264679		-0.6819130		-0.47940912
[4,]		-0.9264679		-0.6819130		-0.47940912
[5,]		0.5388507		-0.6819130		-0.47940912
[6,]		1.5855083		-0.6819130		-0.47940912

	PURCHASES_TRX	CREDIT_LIMIT	PAYMENTS	MINIMUM_PAYMENTS	PRC_FULL_PAYMENT	TENURE
[1,]	-0.5175930	-0.9625197	-0.5439104	-0.30548994	-0.5376958	0.3551601
[2,]	-0.5970196	0.6771649	0.7968061	0.08768365	0.2123677	0.3551601
[3,]	-0.1204598	0.8138052	-0.3994801	-0.09990033	-0.5376958	0.3551601
[4,]	-0.5573063	-0.9078636	-0.3801428	-0.26111544	-0.5376958	0.3551601
[5,]	-0.2793130	-0.7438951	-0.1321118	0.65032579	-0.5376958	0.3551601
[6,]	1.9446329	2.4534898	1.5704929	-0.28080945	2.8375934	0.3551601

Uji Kelayakan Data

Sebelum dilakukan analisis faktor lebih lanjut pada data, langkah pertama yang harus dilakukan adalah menguji apakah data yang dimiliki layak untuk dilakukan analisis faktor atau tidak dengan menggunakan uji Bartlett (*Bartlett's Test of Sphericity*) dan KMO (Kaiser-Meyer-Olkin) *Measure of Sampling Adequacy*.

Uji Bartlett digunakan untuk mengetahui adanya korelasi antar variabel, sedangkan KMO digunakan untuk mengukur kecukupan sampling. Apabila nilai MSA pada uji KMO lebih besar dari 0,5 dan p-value yang dihasilkan pada uji Bartlett kurang dari taraf signifikansi (α) 0,05 maka data dapat dikatakan layak untuk digunakan pada analisis faktor

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Dengan menggunakan bantuan *software* Rstudio diperoleh

```
> ### Uji Kelayakan Data ###
> # Mengukur Kecukupan Sampling
> #Uji KMO (Kaiser Meyer Olkin)
> KMO(scaled.data)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = scaled.data)
Overall MSA = 0.64
MSA for each item =
```

BALANCE	BALANCE_FREQUENCY
0.71	0.75
PURCHASES	ONEOFF_PURCHASES
0.59	0.51
INSTALLMENTS_PURCHASES	CASH_ADVANCE
0.47	0.78
PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FREQUENCY
0.60	0.53
PURCHASES_INSTALLMENTS_FREQUENCY	CASH_ADVANCE_FREQUENCY
0.54	0.75
CASH_ADVANCE_TRX	PURCHASES_TRX
0.73	0.87
CREDIT_LIMIT	PAYMENTS
0.75	0.85
MINIMUM_PAYMENTS	PRC_FULL_PAYMENT
0.66	0.78
TENURE	
0.48	

```
> # Mengetahui adanya korelasi pada variabel
> #Uji Bartlett
> library(REdaS)
> library(psych)
> bart_spher(scaled.data, use = "everything")

Bartlett's Test of Sphericity

Call: bart_spher(x = scaled.data, use = "everything")

X2 = 168066.75
df = 136
p-value < 2.22e-16
```

Berdasarkan hasil yang diperoleh dengan bantuan Rstudio, untuk uji Bartlett diperoleh $p - value < 2.22e - 16 < 0.05 = \alpha$ dan untuk KMO diperoleh $MSA = 0.64 > 0.5$ sehingga dapat dilakukan analisis faktor pada data

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



• Exploratory Factor Analysis

Matriks Korelasi

$$R = D_s^{-1}SD_s^{-1}$$

Untuk mencari matriks korelasi dapat digunakan function R corr() pada dataframe::

```
> ### Matriks Korelasi ###
> correlation.matrix <- rcorr(as.matrix(df1))
> round(correlation.matrix$r,3)
```

	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES
BALANCE	1.000	0.310	0.176	0.160
BALANCE_FREQUENCY	0.310	1.000	0.123	0.095
PURCHASES	0.176	0.123	1.000	0.917
ONEOFF_PURCHASES	0.160	0.095	0.917	1.000
INSTALLMENTS_PURCHASES	0.122	0.115	0.679	0.330
CASH_ADVANCE	0.496	0.089	-0.054	-0.033
PURCHASES_FREQUENCY	-0.088	0.228	0.393	0.265
ONEOFF_PURCHASES_FREQUENCY	0.064	0.187	0.497	0.525
PURCHASES_INSTALLMENTS_FREQUENCY	-0.070	0.184	0.316	0.128
CASH_ADVANCE_FREQUENCY	0.445	0.181	-0.125	-0.086
CASH_ADVANCE_TRX	0.382	0.133	-0.070	-0.049
PURCHASES_TRX	0.148	0.183	0.689	0.545
CREDIT_LIMIT	0.536	0.088	0.358	0.321
PAYMENTS	0.323	0.039	0.607	0.571
MINIMUM_PAYMENTS	0.399	0.133	0.094	0.049
PRC_FULL_PAYMENT	-0.334	-0.157	0.176	0.130
TENURE	0.067	0.105	0.085	0.063
	INSTALLMENTS_PURCHASES	CASH_ADVANCE		
BALANCE		0.122	0.496	
BALANCE_FREQUENCY		0.115	0.089	
PURCHASES		0.679	-0.054	
ONEOFF_PURCHASES		0.330	-0.033	
INSTALLMENTS_PURCHASES		1.000	-0.066	
CASH_ADVANCE		-0.066	1.000	
PURCHASES_FREQUENCY		0.441	-0.219	
ONEOFF_PURCHASES_FREQUENCY		0.212	-0.090	
PURCHASES_INSTALLMENTS_FREQUENCY		0.511	-0.179	
CASH_ADVANCE_FREQUENCY		-0.136	0.629	
CASH_ADVANCE_TRX		-0.077	0.657	
PURCHASES_TRX		0.626	-0.078	
CREDIT_LIMIT		0.258	0.304	
PAYMENTS		0.385	0.459	
MINIMUM_PAYMENTS		0.132	0.140	
PRC_FULL_PAYMENT		0.178	-0.159	
TENURE		0.083	-0.073	
	PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FREQUENCY		
BALANCE		-0.088	0.064	
BALANCE_FREQUENCY		0.228	0.187	
PURCHASES		0.393	0.497	
ONEOFF_PURCHASES		0.265	0.525	
INSTALLMENTS_PURCHASES		0.441	0.212	
CASH_ADVANCE		-0.219	-0.090	
PURCHASES_FREQUENCY		1.000	0.502	
ONEOFF_PURCHASES_FREQUENCY		0.502	1.000	
PURCHASES_INSTALLMENTS_FREQUENCY		0.862	0.144	
CASH_ADVANCE_FREQUENCY		-0.317	-0.118	
CASH_ADVANCE_TRX		-0.209	-0.073	
PURCHASES_TRX		0.567	0.544	
CREDIT_LIMIT		0.121	0.295	
PAYMENTS		0.101	0.244	
MINIMUM_PAYMENTS		0.003	-0.030	
PRC_FULL_PAYMENT		0.305	0.154	
TENURE		0.052	0.080	
	PURCHASES_INSTALLMENTS_FREQUENCY			
BALANCE		-0.070		
BALANCE_FREQUENCY		0.184		

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



PURCHASES	0.316		
ONEOFF_PURCHASES	0.128		
INSTALLMENTS_PURCHASES	0.511		
CASH_ADVANCE	-0.179		
PURCHASES_FREQUENCY	0.862		
ONEOFF_PURCHASES_FREQUENCY	0.144		
PURCHASES_INSTALLMENTS_FREQUENCY	1.000		
CASH_ADVANCE_FREQUENCY	-0.269		
CASH_ADVANCE_TRX	-0.173		
PURCHASES_TRX	0.529		
CREDIT_LIMIT	0.063		
PAYMENTS	0.083		
MINIMUM_PAYMENTS	0.030		
PRC_FULL_PAYMENT	0.250		
TENURE	0.066		
	CASH_ADVANCE_FREQUENCY	CASH_ADVANCE_TRX	PURCHASES_TRX
BALANCE	0.445	0.382	0.148
BALANCE_FREQUENCY	0.181	0.133	0.183
PURCHASES	-0.125	-0.070	0.689
ONEOFF_PURCHASES	-0.086	-0.049	0.545
INSTALLMENTS_PURCHASES	-0.136	-0.077	0.626
CASH_ADVANCE	0.629	0.657	-0.078
PURCHASES_FREQUENCY	-0.317	-0.209	0.567
ONEOFF_PURCHASES_FREQUENCY	-0.118	-0.073	0.544
PURCHASES_INSTALLMENTS_FREQUENCY	-0.269	-0.173	0.529
CASH_ADVANCE_FREQUENCY	1.000	0.800	-0.137
CASH_ADVANCE_TRX	0.800	1.000	-0.070
PURCHASES_TRX	-0.137	-0.070	1.000
CREDIT_LIMIT	0.133	0.150	0.274
PAYMENTS	0.182	0.256	0.371
MINIMUM_PAYMENTS	0.099	0.110	0.096
PRC_FULL_PAYMENT	-0.260	-0.177	0.157
TENURE	-0.140	-0.045	0.119
	CREDIT_LIMIT	PAYMENTS	MINIMUM_PAYMENTS
BALANCE	0.536	0.323	0.399
BALANCE_FREQUENCY	0.088	0.039	0.133
PURCHASES	0.358	0.607	0.094
ONEOFF_PURCHASES	0.321	0.571	0.049
INSTALLMENTS_PURCHASES	0.258	0.385	0.132
CASH_ADVANCE	0.304	0.459	0.140
PURCHASES_FREQUENCY	0.121	0.101	0.003
ONEOFF_PURCHASES_FREQUENCY	0.295	0.244	-0.030
PURCHASES_INSTALLMENTS_FREQUENCY	0.063	0.083	0.030
CASH_ADVANCE_FREQUENCY	0.133	0.182	0.099
CASH_ADVANCE_TRX	0.150	0.256	0.110
PURCHASES_TRX	0.274	0.371	0.096
CREDIT_LIMIT	1.000	0.427	0.127
PAYMENTS	0.427	1.000	0.127
MINIMUM_PAYMENTS	0.127	0.127	1.000
PRC_FULL_PAYMENT	0.053	0.105	-0.140
TENURE	0.136	0.103	0.059
	PRC_FULL_PAYMENT	TENURE	
BALANCE	-0.334	0.067	
BALANCE_FREQUENCY	-0.157	0.105	
PURCHASES	0.176	0.085	
ONEOFF_PURCHASES	0.130	0.063	
INSTALLMENTS_PURCHASES	0.178	0.083	
CASH_ADVANCE	-0.159	-0.073	
PURCHASES_FREQUENCY	0.305	0.052	
ONEOFF_PURCHASES_FREQUENCY	0.154	0.080	
PURCHASES_INSTALLMENTS_FREQUENCY	0.250	0.066	
CASH_ADVANCE_FREQUENCY	-0.260	-0.140	
CASH_ADVANCE_TRX	-0.177	-0.045	
PURCHASES_TRX	0.157	0.119	
CREDIT_LIMIT	0.053	0.136	
PAYMENTS	0.105	0.103	
MINIMUM_PAYMENTS	-0.140	0.059	
PRC_FULL_PAYMENT	1.000	-0.024	
TENURE	-0.024	1.000	

Terlihat bahwa antar variabel berkorelasi sangat tinggi. Dari hasil matriks korelasi, kami menduga bahwa faktor yang akan terbentuk sebanyak 6 faktor. Selanjutnya, kami membuat *scree plot* untuk menentukan banyaknya faktor.

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534

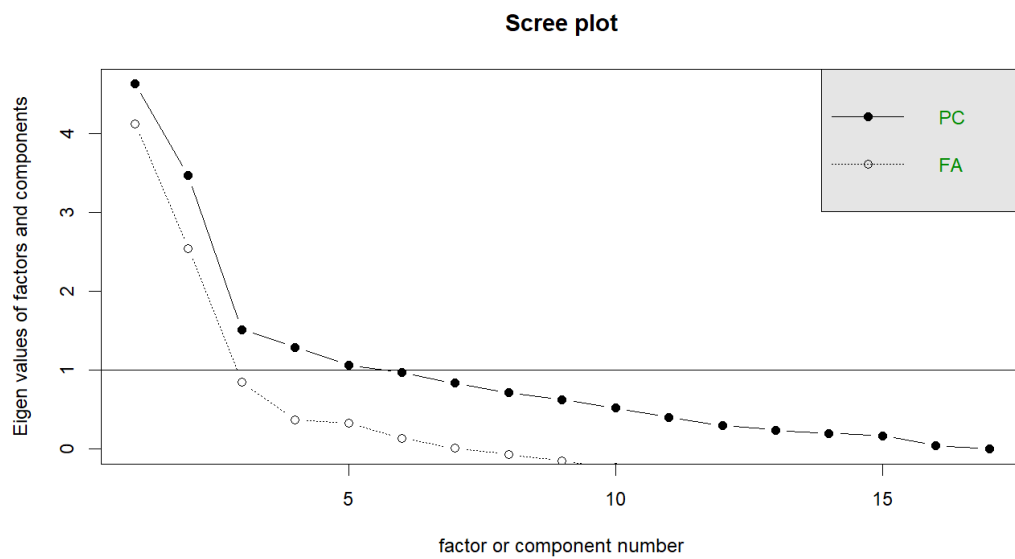


Menentukan Banyaknya Faktor

➤ *Scree Plot*

Dalam menentukan banyaknya faktor yang akan digunakan dalam analisis faktor, salah satu cara yang dapat dilakukan adalah dengan melihat *scree plot* dari data.

Dengan menggunakan *software* Rstudio diperoleh:



Terdapat 2 cara untuk menentukan banyaknya faktor berdasarkan *scree plot*, yaitu:

1. Melihat nilai eigen yang lebih dari 1
2. Berdasarkan titik sebelum garis lurus (tidak ada penurunan yang signifikan)

UAS Analisis Data

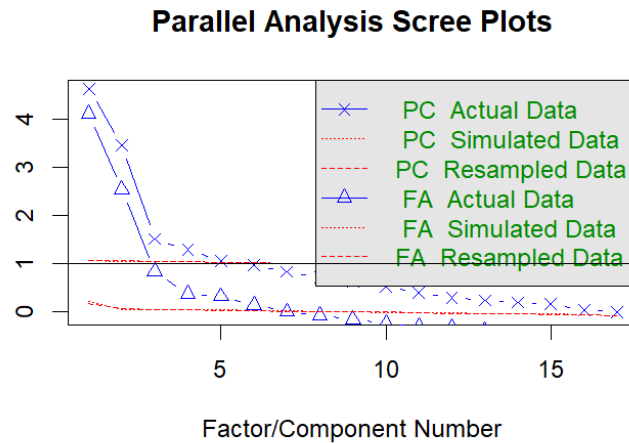
Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



➤ Parallel Analysis

jenvalues of principal components and factor an:



Dengan menggunakan *software* Rstudio diperoleh:

```
> x <- fa.parallel(df1, fm="pa", fa="both", n.iter=1)
Parallel analysis suggests that the number of factors = 6 and the number of
components = 5
```

∴ Jadi, diperoleh bahwa banyaknya faktor yang akan diambil adalah 6.

Maka dari itu akan dilakukan EFA untuk 6 faktor. Untuk estimasi *factor loadings*, akan digunakan *factor analysis method* dengan rotasi varimax.

Factor Loadings (Facto Analysis Model)

$$\begin{aligned}\hat{\Lambda} &= \mathbf{C}_1 \mathbf{D}_1^{\frac{1}{2}} \\ \begin{bmatrix} \hat{\lambda}_{11} & \hat{\lambda}_{12} \\ \hat{\lambda}_{21} & \hat{\lambda}_{22} \\ \vdots & \vdots \\ \hat{\lambda}_{18\ 1} & \hat{\lambda}_{18\ 2} \end{bmatrix} &= \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ \vdots & \vdots \\ c_{18\ 1} & c_{18\ 2} \end{bmatrix} \begin{bmatrix} \sqrt{\theta_1} & 0 \\ 0 & \sqrt{\theta_2} \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{\theta_1} c_{11} & \sqrt{\theta_2} c_{12} \\ \sqrt{\theta_1} c_{21} & \sqrt{\theta_2} c_{22} \\ \vdots & \vdots \\ \sqrt{\theta_1} c_{18\ 1} & \sqrt{\theta_2} c_{18\ 2} \end{bmatrix}\end{aligned}$$

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Varimax Rotation

$$\hat{\Lambda}^* = \hat{\Lambda} T$$

Untuk mencari nilai *factor loadings* dapat digunakan *package* “psych”:

```
> ### Menentukan Analisis Faktor ###
> fit <- fa(dfl, x$nfact, rotate="promax", fm="pa")
> print(fit$loadings, cutoff=.3)
```

Loadings:

	PA1	PA2	PA3	PA4	PA5	PA6
BALANCE				0.971		
BALANCE_FREQUENCY						-0.363
PURCHASES	1.131					
ONEOFF_PURCHASES	0.849					
INSTALLMENTS_PURCHASES	0.457		0.428			
CASH_ADVANCE		0.745				
PURCHASES_FREQUENCY			0.897			
ONEOFF_PURCHASES_FREQUENCY					0.939	
PURCHASES_INSTALLMENTS_FREQUENCY			1.076			
CASH_ADVANCE_FREQUENCY		0.894				
CASH_ADVANCE_TRX		0.886				
PURCHASES_TRX	0.440		0.373			
CREDIT_LIMIT				0.460		
PAYMENTS	0.480	0.323				0.425
MINIMUM_PAYMENTS				0.423		
PRC_FULL_PAYMENT				-0.312		0.386
TENURE						

	PA1	PA2	PA3	PA4	PA5	PA6
SS loadings	2.726	2.341	2.476	1.637	1.235	0.729
Proportion Var	0.160	0.138	0.146	0.096	0.073	0.043
Cumulative Var	0.160	0.298	0.444	0.540	0.613	0.656

Communalities

$$\hat{h}_i^2 = \sum_{j=1}^2 \hat{\lambda}_{ij}^2$$

Untuk mencari nilai *communalities* dapat digunakan *package* “psych”:

```
> ### Communalities ###
> efa$communality
```

	BALANCE	BALANCE_FREQUENCY
	0.7500325	0.6813407
	PURCHASES	ONEOFF_PURCHASES
	0.9267827	0.8695967
INSTALLMENTS_PURCHASES		CASH_ADVANCE
	0.6920171	0.7849320
PURCHASES_FREQUENCY		ONEOFF_PURCHASES_FREQUENCY
	0.8762159	0.7229230
PURCHASES_INSTALLMENTS_FREQUENCY		CASH_ADVANCE_FREQUENCY
	0.8927027	0.8173299
CASH_ADVANCE_TRX		PURCHASES_TRX
	0.7859875	0.7277429
CREDIT_LIMIT		PAYMENTS
	0.5569473	0.7440125
MINIMUM_PAYMENTS		PRC_FULL_PAYMENT
	0.6682010	0.6122147

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



TENURE
0.8257821

Specific Variances

$$\hat{\psi}_I = 1 - \hat{h}_i^2$$

Untuk mencari nilai *specific variances* dapat digunakan *package* “psych”:

```
> ### Specific Variances ###
> efa$uniquenesses
```

BALANCE	BALANCE_FREQUENCY
0.24996750	0.31865928
PURCHASES	ONEOFF_PURCHASES
0.07321729	0.13040326
INSTALLMENTS_PURCHASES	CASH_ADVANCE
0.30798293	0.21506797
PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FREQUENCY
0.12378414	0.27707704
PURCHASES_INSTALLMENTS_FREQUENCY	CASH_ADVANCE_FREQUENCY
0.10729731	0.18267009
CASH_ADVANCE_TRX	PURCHASES_TRX
0.21401246	0.27225706
CREDIT_LIMIT	PAYMENTS
0.44305270	0.25598748
MINIMUM_PAYMENTS	PRC_FULL_PAYMENT
0.33179901	0.38778533
TENURE	
0.17421786	

Summary Model

Dengan menggunakan *software* Rstudio diperoleh:

```
> summary(efa)
```

Factor analysis with Call: principal(r = df1, nfactors = 6, rotate = "varimax")

Test of the hypothesis that 6 factors are sufficient.
The degrees of freedom for the model is 49 and the objective function was 11.42
The number of observations was 8636 with Chi Square = 98528.64 with prob
< 0

The root mean square of the residuals (RMSA) is 0.06

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

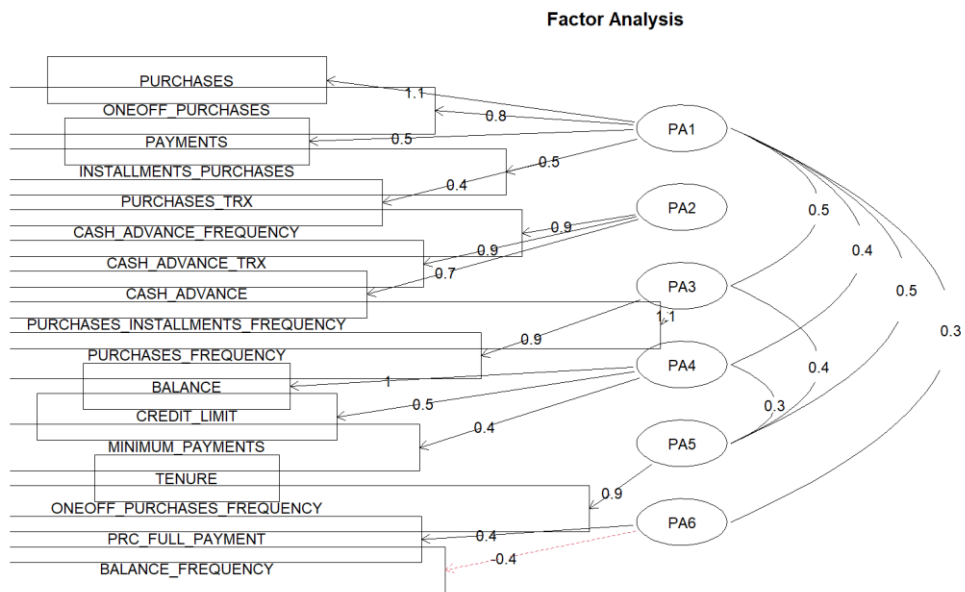
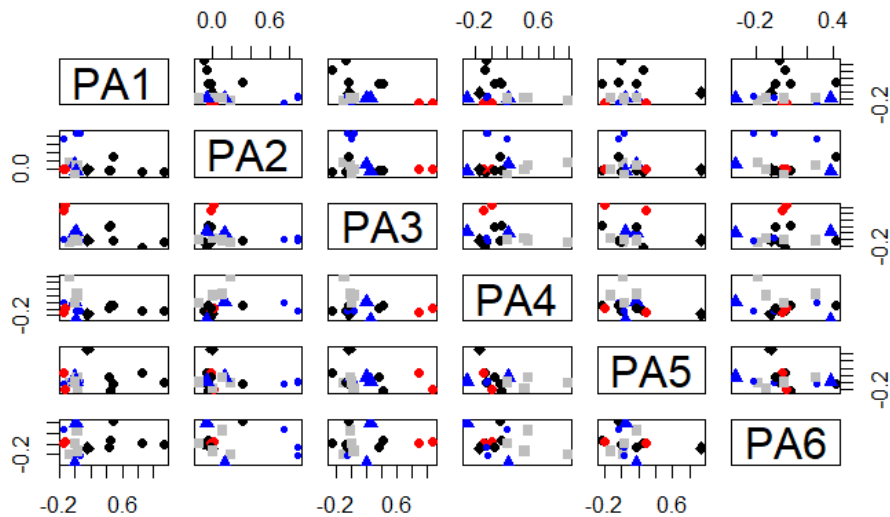
NPM : 1906299534



Diagram Factor Analysis

Dengan menggunakan *software* Rstudio diperoleh:

Factor Analysis



UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



∴ Jadi, berdasarkan *output factor loadings* dan diagram *factor analysis* diperoleh 6 faktor diantaranya

Faktor 1 : PURCHASES, ONEOFF_PURCHASES, PAYMENTS,
INSTALLMENTS_ PURCHASES, PURCHASES_TRX

Faktor 2 : CASH_ADVANCE_FREQUENCY, CASH_ADVANCE_TRX,
CASH_ADVANCE

Faktor 3 : PURCHASES_INSTALLMENTS_FREQUENCY,
PURCHASES_FREQUENCY

Faktor 4 : BALANCE, CREDIT_LIMIT, MINIMUM_PAYMENTS

Faktor 5 : TENURE, ONEOFF_PURCHASES_FREQUENCY

Faktor 6 : PRC_FULL_PAYMENT, BALANCE_FREQUENCY

Model tersebut dapat menjelaskan 65.6% variansi dari data.

3. *Clustering Analysis (CA)*

Pada analisis cluster, metode yang akan digunakan adalah metode *K-Means* (non-hirarki) karena observasi yang cukup banyak pada data set yang tersedia (data segmentasi klien).

Tujuan:

Untuk mengklasifikasi objek atau kasus (responden) ke dalam kelompok yang relatif homogen.

Langkah Kerja:

1. Merumuskan Masalah

Dalam merumuskan masalah, hal yang terpenting adalah memilih dan menentukan variabel-variabel yang akan digunakan untuk membentuk cluster.

2. Memilih Ukuran Jarak dan Similaritas

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Objek dengan jarak yang lebih pendek antara mereka akan lebih mirip atau sama lain dibandingkan dengan pasangan dengan jarak yang lebih panjang.

3. Standarisasi Data

Proses standarisasi dilakukan apabila diantara variabel-variabel yang diteliti terdapat perbedaan ukuran satuan yang besar karena perbedaan satuan yang mencolok mengakibatkan perhitungan analisis cluster menjadi tidak valid.

4. Memilih Suatu Prosedur Pengklasteran

Menentukan metode cluster dengan metode hirarki atau non-hirarki

5. Melakukan Interpretasi terhadap Cluster yang Telah Terbentuk

Tahap inteprestasi meliputi pengujian tiap cluster dalam terminologi untuk menamai dan menandai dengan suatu label yang secara akurat dapat menjelaskan kealamian cluster. Proses ini dimulai dengan suatu ukuran yang sering digunakan, yaitu Centroid Cluster

6. Melakukan Validasi dan Profiling Cluster

a. Proses Validasi Solusi Cluster

Bertujuan menjamin bahwa solusi yang dihasilkan dari Analisis Cluster dapat mewakili populasi dan dapat digenerelisasi untuk objek lain.

b. Pembuatan Profil

Untuk menjelaskan karakteristik setiap cluster berdasar profil tertentu.

Standardisasi Data

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_i}{s_i}, \quad i = 1, 2, \dots, 18$$

Standarisasi data dilakukan untuk mengubah nilai asli data menjadi bentuk distribusi normal. Standarisasi data dapat menggunakan function R `scale()` pada *dataframe*:

```
> ### Standarisasi Data ###
> scaled.data <- scale(df1)
> head(scaled.data)
      BALANCE BALANCE_FREQUENCY PURCHASES ONEOFF_PURCHASES INSTALLMENTS_PURCHASES
[1,] -0.7445817      -0.37002536  -0.4291590      -0.35913949      -0.3548054
[2,]  0.7641079       0.06767501  -0.4731808      -0.35913949      -0.4588125
[3,]  0.4265777       0.50537539  -0.1164058       0.09990369      -0.4588125
[4,] -0.3738888       0.50537539  -0.4657977      -0.34964003      -0.4588125
[5,]  0.0995451       0.50537539  0.1420539      -0.35913949       0.9947574
[6,] -0.4647726       0.50537539  2.7989266       3.44220252       0.2916739
```

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



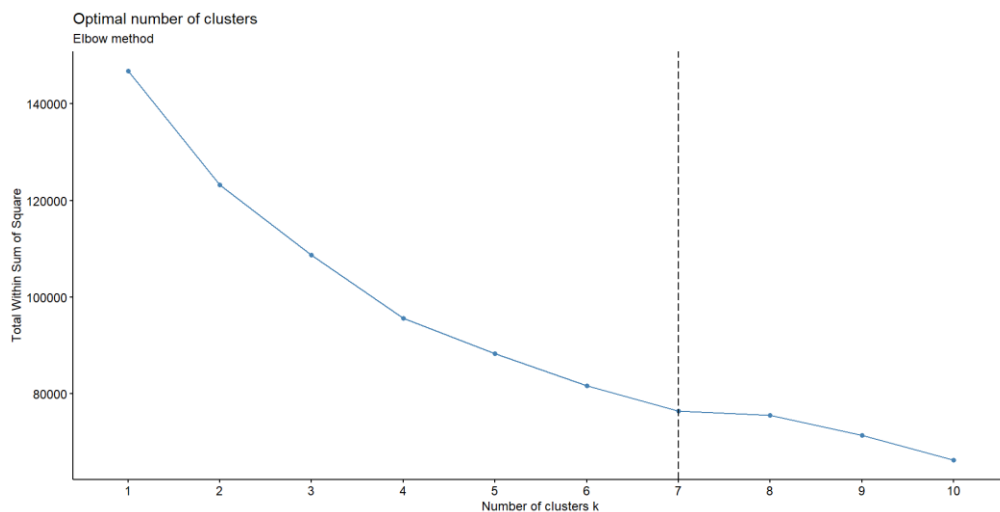
	CASH_ADVANCE	PURCHASES	FREQUENCY	ONEOFF_PURCHASES	FREQUENCY	
[1,]	-0.4686284		-0.8207213		-0.6862398	
[2,]	2.5684078		-1.2360673		-0.6862398	
[3,]	-0.4686284		1.2560039		2.6464980	
[4,]	-0.4686284		-1.0283956		-0.4085128	
[5,]	-0.4686284		0.4253143		-0.6862398	
[6,]	-0.4686284		1.2560039		2.6464980	
	PURCHASES_INSTALLMENTS	FREQUENCY	CASH_ADVANCE	FREQUENCY	CASH_ADVANCE_TRX	
[1,]		-0.7171374		-0.6819130	-0.47940912	
[2,]		-0.9264679		0.5569899	0.09925221	
[3,]		-0.9264679		-0.6819130	-0.47940912	
[4,]		-0.9264679		-0.6819130	-0.47940912	
[5,]		0.5388507		-0.6819130	-0.47940912	
[6,]		1.5855083		-0.6819130	-0.47940912	
	PURCHASES_TRX	CREDIT_LIMIT	PAYMENTS	MINIMUM_PAYMENTS	PRC_FULL_PAYMENT	TENURE
[1,]	-0.5175930	-0.9625197	-0.5439104	-0.30548994	-0.5376958	0.3551601
[2,]	-0.5970196	0.6771649	0.7968061	0.08768365	0.2123677	0.3551601
[3,]	-0.1204598	0.8138052	-0.3994801	-0.09990033	-0.5376958	0.3551601
[4,]	-0.5573063	-0.9078636	-0.3801428	-0.26111544	-0.5376958	0.3551601
[5,]	-0.2793130	-0.7438951	-0.1321118	0.65032579	-0.5376958	0.3551601
[6,]	1.9446329	2.4534898	1.5704929	-0.28080945	2.8375934	0.3551601

Jumlah Kluster Optimal

Untuk mencari jumlah optimal kluster akan digunakan tiga metode, yaitu *Elbow Method*, *Silhouette Square Method*, dan *Gap Stat Method*.

- Metode Elbow

```
> # Elbow method
> fviz_nbclust(scaled.data, kmeans, method = "wss") +
+   geom_vline(xintercept = 7, linetype = 5)+
+   labs(subtitle = "Elbow method")
```



UAS Analisis Data

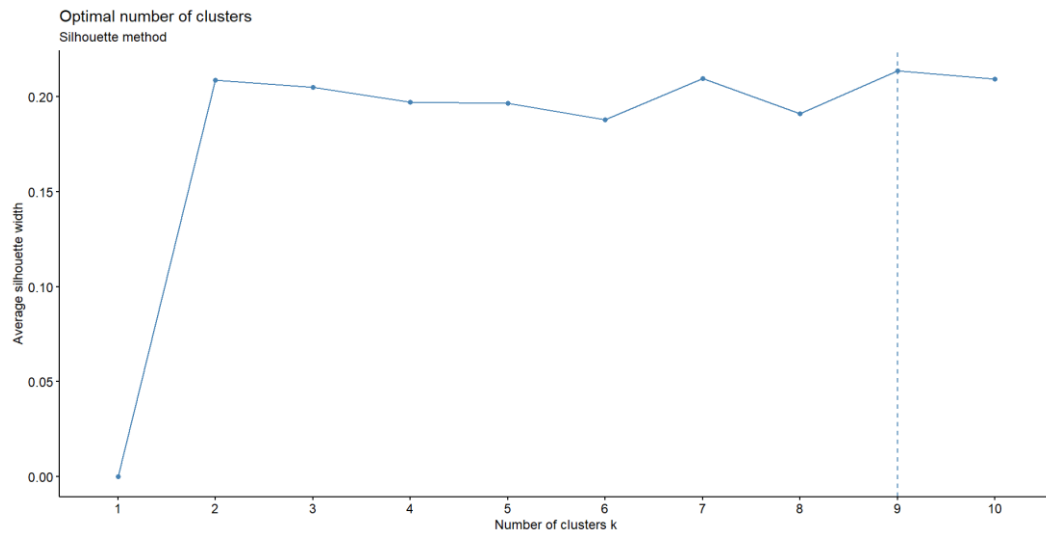
Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



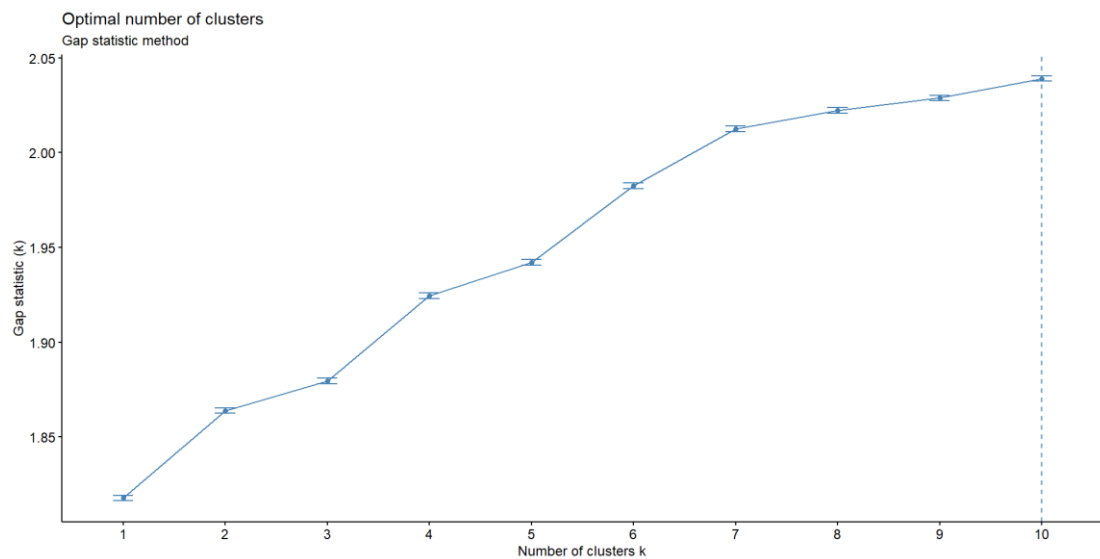
- Metode *Silhouette Square*

```
> # Silhouette method
> fviz_nbclust(scaled.data, kmeans, method = "silhouette")+
+   labs(subtitle = "Silhouette method")
```



- Metode *Gap Stat*

```
> # Gap statistic
> set.seed(123)
> fviz_nbclust(scaled.data, kmeans, nstart = 25, method = "gap_stat", nboot
= 50)+
+   labs(subtitle = "Gap statistic method")
```



UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Dapat terlihat bahwa dari ketiga plot di atas, terdapat tiga hasil yang berbeda untuk jumlah optimal klaster yang akan dibentuk, yaitu 7 klaster, 9 klaster, dan 10 klaster.

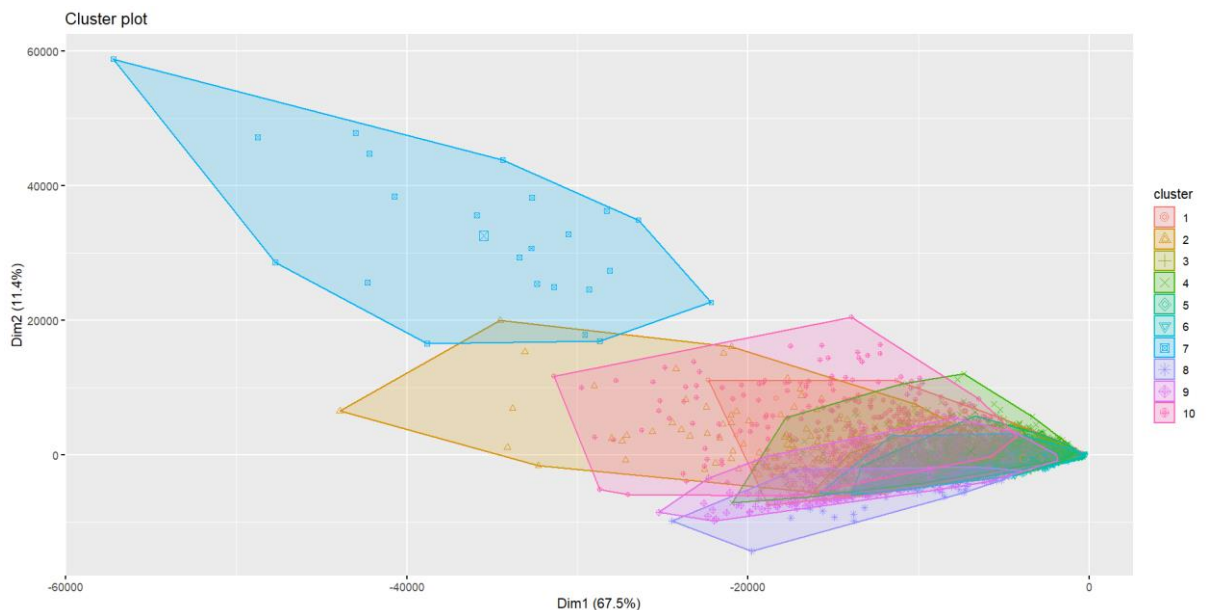
Pembentukan Klaster

Karena terdapat dua hasil yang berbeda, maka akan dibandingkan kedua hasil tersebut dengan membandingkan nilai dari *betweens* dan jumlah *withinss*.

```
> ### Pembentukan Klaster ###
> #Kmeans clustering
> klaster7 <- kmeans(scaled.data, centers=7, nstart=25)
> klaster9 <- kmeans(scaled.data, centers=9, nstart=25)
> klaster10 <- kmeans(scaled.data, centers=10, nstart=25)
> data.frame(klaster7$tot.withinss, klaster9$tot.withinss,
klaster10$tot.withinss)
  klaster7.tot.withinss klaster9.tot.withinss klaster10.tot.withinss
1          76430.29          67062.86          63776.85
> data.frame(klaster7$betweenss, klaster9$betweenss, klaster10$betweenss)
  klaster7.betweenss klaster9.betweenss klaster10.betweenss
1          70364.71          79732.14          83018.15
```

Karena klaster 10 memiliki nilai *betweens* terbesar dan jumlah *withinss* terkecil maka pembentukan jumlah optimal adalah 10 klaster

Dengan bantuan *software* Rstudio diperoleh *Cluster Plot* sebagai berikut:



UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Perbandingan Hasil Tiap Metode

Untuk mendapatkan informasi mengenai segmentasi klien akan dilakukan analisis data dengan menggunakan metode *principal component analysis* (PCA), *factor analysis* (FA), dan *clustering analysis* (CA).

Principal Component Analysis (PCA) digunakan untuk mereduksi variabel-variabel sehingga mengurangi kompleksitas hubungan timbal-balik antara sejumlah besar variabel yang diamati ke sejumlah relatif kecil dari kombinasi linearnya, yang disebut sebagai komponen utama. Dengan menggunakan metode *principal component analysis* (PCA) diperoleh 5 komponen utama dengan model dapat menjelaskan 70.37% variansi dari data.

Factor Analysis (FA) digunakan untuk mengidentifikasi hubungan antar variabel dengan menggunakan teknik *Exploratory Factor Analysis*, serta akan dilakukan analisis apakah data dapat direduksi ke dalam beberapa faktor dengan menggunakan *Confirmatory Factor Analysis*. Untuk mereduksi data dengan mencari kombinasi linear dari variabel tersembunyi yang tidak diketahui atau faktor yang ingin dipelajari sekaligus mengidentifikasi apakah terdapat korelasi antar variabel. Hal tersebut dapat dimulai dengan mengasumsikan bahwa terdapat variabel yang membentuk variabel-variabel yang diukur dalam data. Dengan menggunakan metode *factor analysis* (FA) diperoleh 6 faktor dengan model dapat menjelaskan 65.6% variansi dari data.

Cluster Analysis (CA) digunakan untuk mengklasifikasi objek atau kasus (responden) ke dalam kelompok yang relatif homogen. Dengan menggunakan metode *clustering analysis* (CA) akan dicari jumlah optimal klaster akan digunakan tiga metode, yaitu *Elbow Method*, *Silhouette Square Method*, dan *Gap Stat Method*. Berdasarkan nilai *between*s terbesar dan jumlah *within*s terkecil maka pembentukan jumlah optimal adalah 10 klaster. Jadi, diperoleh 10 klaster

UAS Analisis Data

Nama : Michael Mario Bramanthyo Adhi

NPM : 1906299534



Rangkuman Kesimpulan

Berdasarkan hasil analisis data yang telah dilakukan, dapat diketahui bahwa untuk mempelajari dan mengetahui segmentasi klien kartu kredit dan mempelajari pola konsumsi klien dapat menggunakan metode *principal component analysis* (PCA), *factor analysis* (FA), dan *clustering analysis* (CA). Segmentasi klien adalah strategi pemasaran yang mengelompokkan target pasar berdasarkan karakteristik yang sama untuk dikelola secara efektif dan tepat agar mencapai tujuan bisnis yang telah ditetapkan. Tujuan segmentasi klien termasuk membagi klien target menjadi kelompok-kelompok yang lebih kecil yang mencerminkan kesamaan di antara klien di setiap kelompok untuk mengembangkan hubungan yang lebih baik dengan memahami kebutuhan setiap segmen klien, mengidentifikasi klien yang berharga, mengidentifikasi peluang *cross-selling* dan *up-selling*, dan meningkatkan profitabilitas dengan mengembangkan strategi pemasaran yang lebih efektif untuk setiap segmen. Dengan menggunakan metode *principal component analysis* (PCA) dapat dilakukan segmentasi menjadi 5 komponen utama. Dengan menggunakan metode *factor analysis* (FA) dapat dilakukan segmentasi menjadi 6 faktor. Selain itu, dengan menggunakan metode *clustering analysis* (CA) dapat dilakukan segmentasi menjadi 10 klaster.