



# **Aplicações pragmáticas de algoritmos de clusterização em instituições financeiras e bancos de varejo**

Aluno: Msc. Osvaldo Luiz dos Santos Pereira

Orientador: Prof. Dr. Thiago Gentil Ramires

- Aplicação do algoritmo K-Means no mercado bancário.
- Grande dimensionalidade no mercado de banco de varejos.
- Milhares de agências bancárias na rede.
- Incapacidade de lidar com a alta dimensionalidade.
- Aglomeração de agências em grupos comparáveis.
- Estratégias gerais entre grupos de agências comparáveis.

- Método de quantização vetorial. Originalmente desenvolvido na década de 1960 para processamento de sinais por MacQueen, 1967. Processamento de sinais.
- Particiona o dataset em K aglomerados (clusters) com as distâncias mais próximas dos centroides. Minimiza variância dentro dos aglomerados. Maximiza a variância fora dos aglomerados. Distâncias quadráticas Euclidianas como métrica.
- Problema computacionalmente complexo (NP-Hard) com alta dimensionalidade.
- Modelo de machine learning não supervisionado.

# Algoritmo K-Means

Dado um conjunto de dados com observações ( $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ ) sendo cada observação um vetor d-dimensional, k-means aglomera  $n$  observações em  $k$  partições no conjunto  $\mathbf{S} = \{S_1, S_2, \dots, S_k\}$  minimizando a soma dos quadrados das distâncias dentro de cada aglomerado (*Within Cluster Sum of Squares*, WCSS do inglês)

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \arg \min_{\mathbf{S}} \sum_{i=1}^k |S_i| \text{Var } S_i \quad (1)$$

Sendo  $\boldsymbol{\mu}_i$  a média dos pontos de cada subconjunto  $S_i$ , ou seja

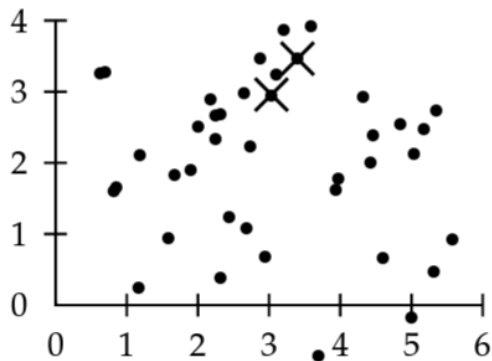
$$\boldsymbol{\mu}_i = \frac{1}{|S_i|} \sum_{\mathbf{x} \in S_i} \mathbf{x}, \quad (2)$$

Sendo  $|S_i|$  o tamanho de  $S_i$ , e  $\|\cdot\|$  é a norma  $L^2$ . É equivalente a minimizar a variância quadrática pareada dos pontos no mesmo aglomerado. De forma equivalente, podemos afirmar matematicamente:

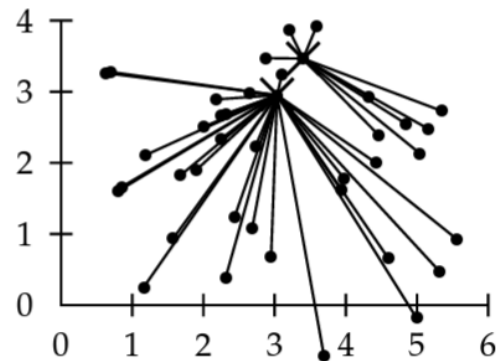
$$|S_i| \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \frac{1}{2} \sum_{\mathbf{x}, \mathbf{y} \in S_i} \|\mathbf{x} - \mathbf{y}\|^2 \quad \rightarrow \quad \arg \min_{\mathbf{S}} \sum_{i=1}^k \frac{1}{|S_i|} \sum_{\mathbf{x}, \mathbf{y} \in S_i} \|\mathbf{x} - \mathbf{y}\|^2 \quad (3)$$

Como a variância total é constante, é equivalente a maximizar a soma do quadrado das distâncias (desvios) entre diferentes pontos dentro do mesmo aglomerado (*Between Clusters Sum of Squares*, BCSS do inglês).

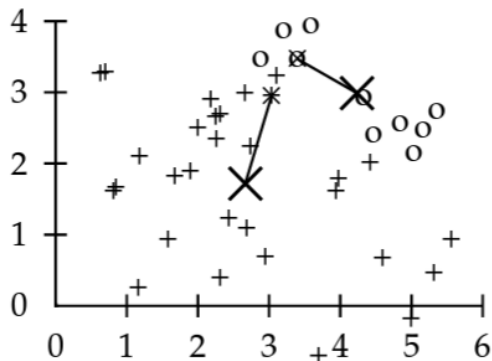
# Algoritmo K-Means



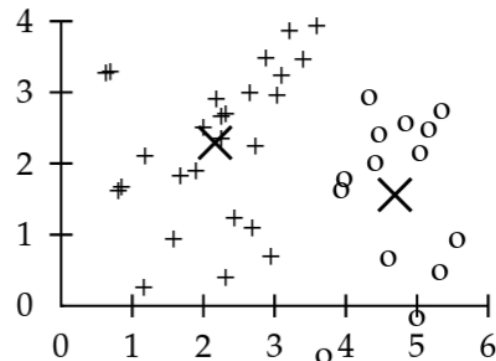
selection of seeds



assignment of documents (iter. 1)



recomputation/movement of  $\vec{\mu}'$ 's (iter. 1)



$\vec{\mu}$ 's after convergence (iter. 9)

K-MEANS( $\{\vec{x}_1, \dots, \vec{x}_N\}, K$ )

```

1   $(\vec{s}_1, \vec{s}_2, \dots, \vec{s}_K) \leftarrow \text{SELECTRANDOMSEEDS}(\{\vec{x}_1, \dots, \vec{x}_N\}, K)$ 
2  for  $k \leftarrow 1$  to  $K$ 
3  do  $\vec{\mu}_k \leftarrow \vec{s}_k$ 
4  while stopping criterion has not been met
5  do for  $k \leftarrow 1$  to  $K$ 
6      do  $\omega_k \leftarrow \{\}$ 
7      for  $n \leftarrow 1$  to  $N$ 
8          do  $j \leftarrow \arg \min_{j'} |\vec{\mu}_{j'} - \vec{x}_n|$ 
9               $\omega_j \leftarrow \omega_j \cup \{\vec{x}_n\}$  (reassignment of vectors)
10     for  $k \leftarrow 1$  to  $K$ 
11         do  $\vec{\mu}_k \leftarrow \frac{1}{|\omega_k|} \sum_{\vec{x} \in \omega_k} \vec{x}$  (recomputation of centroids)
12 return  $\{\vec{\mu}_1, \dots, \vec{\mu}_K\}$ 

```

► **Figure 16.5** The K-means algorithm. For most IR applications, the vectors  $\vec{x}_n \in \mathbb{R}^M$  should be length-normalized. Alternative methods of seed selection and initialization are discussed on page 364.

- Banco de varejo com 5.200 agências em sua rede.
- Atribuição e apuração de metas (mensal, trimestral, semestral, anual).
- Incapacidade de avaliar 5.200 metas mensalmente.
- Variáveis de performance (características das agências).
- Agrupamento de variáveis: Tamanho, Potencial de negócios, Apetite por crédito, Perfil de investimento, Volume transacionado, Valor médio transacionado.
- “*Benchmark*” das variáveis de agrupamento (*clusterização*).
- Metodologia de comparação entre os resultados médios por grupo homogêneo.

# Variáveis de perfil

Tamanho (Size)	Variáveis de Clusterização
Potencial de Negócios (Business Potential)	<p><b>NumberTellerCapacity:</b> número de caixas humanos da agência</p> <p><b>NumberManagerPersonalCapacity:</b> número de gerentes que atendem pessoa física</p> <p><b>NumberManagerBusinessCapacity:</b> número de gerentes que atendem pessoa jurídica</p> <p><b>NumberATM:</b> número de caixas eletrônicos</p> <p><b>NumberPersonalClientsTierA:</b> número total de clientes pessoa física do tier A</p> <p><b>NumberPersonalClientsTierB:</b> número total de clientes pessoa física do tier B</p> <p><b>NumberPersonalClientsTierC:</b> número total de clientes pessoa física do tier C</p> <p><b>NumberPersonalClientsTierD:</b> número total de clientes pessoa física do tier D</p> <p><b>NumberINSSClients:</b> número de clientes INSS</p> <p><b>NumberSalaryAccounts:</b> número total de clientes com conta salário</p> <p><b>NBusinessClientsTierA:</b> número de contas pessoa jurídica do tier A</p> <p><b>NBusinessClientsTierB:</b> número de contas pessoa jurídica do tier B</p> <p><b>NBusinessClientsTierC:</b> número de contas pessoa jurídica do tier C</p> <p><b>NBusinessClientsTierD:</b> número de contas pessoa jurídica do tier D</p> <p><b>BranchSizeSquareMeters:</b> tamanho da agência em metros quadrados</p>
Crédito (Credit)	
Investimentos (Investment)	
Saúde Financeira (Financial Health)	
Volume de Transações (Transaction Volume)	
Valor Médio de Transações (Transaction Value)	

# Variáveis de perfil

Tamanho (Size)

Potencial de Negócios  
(Business Potential)

Crédito (Credit)

Investimentos  
(Investment)

Saúde Financeira  
(Financial Health)

Volume de Transações  
(Transaction Volume)

Valor Médio de Transações  
(Transaction Value)

## Variáveis de Clusterização

**AvgMonthlyIncomePersonalClientTierA:** valor médio de renda mensal de clientes do tier A

**AvgMonthlyIncomePersonalClientTierB:** valor médio de renda mensal de clientes tier B

**AvgMonthlyIncomePersonalClientTierC:** valor médio de renda mensal de clientes tier C

**AvgMonthlyIncomePersonalClientTierD:** valor médio de renda mensal de clientes tier D

**AvgMonthlyINSSBenefitsAmount:** valor médio do benefício recebido por clientes INSS

**AvgMonthlySalaryValue:** valor médio da renda mensal de clientes pessoa física

**AvgMonthlyEBITDABusinessClientTierA:** faturamento mensal pessoa jurídica do tier A

**AvgMonthlyEBITDABusinessClientTierB:** faturamento mensal pessoa jurídica do tier B

**AvgMonthlyEBITDABusinessClientTierC:** faturamento mensal pessoa jurídica do tier C

**AvgMonthlyEBITDABusinessClientTierD:** faturamento mensal pessoa jurídica do tier D



# Variáveis de perfil

Tamanho (Size)

Potencial de Negócios  
(Business Potential)

Crédito (Credit)

Investimentos  
(Investment)

Saúde Financeira  
(Financial Health)

Volume de Transações  
(Transaction Volume)

Valor Médio de Transações  
(Transaction Value)

## Variáveis de Clusterização

**AvgMonthlyManagerPersonalLoanTransactions:** quantidade média mensal de transações do tipo empréstimo realizadas por gerentes de pessoa física da agência

**AvgMonthlyManagerBusinessLoanTransactions:** quantidade média mensal de transações do tipo empréstimos realizadas por gerentes de pessoa jurídica da agência

**AvgMonthlyPersonalCreditCardPaymentAmount:** valor médio mensal do valor de pagamento de faturas mensais de cartão de crédito por contas corrente pessoa física

**AvgMonthlyBusinessCreditCardPaymentAmount:** valor médio mensal do valor de pagamento de faturas mensais de cartão de crédito por contas corrente pessoa jurídica

# Variáveis de perfil

Tamanho (Size)

Potencial de Negócios  
(Business Potential)

Crédito (Credit)

**Investimentos  
(Investment)**

Saúde Financeira  
(Financial Health)

Volume de Transações  
(Transaction Volume)

Valor Médio de Transações  
(Transaction Value)

## Variáveis de Clusterização

**AvgMonthlyManagerPersonalInvestmentTransactions:** quantidade média mensal de transações do tipo investimento realizadas por gerente de pessoa física da agência

**AvgMonthlyManagerBusinessInvestmentTransactions:** quantidade média mensal de transações do tipo investimento realizadas por gerente pessoa jurídica da agência

Tamanho (Size)

Potencial de Negócios  
(Business Potential)

Crédito (Credit)

Investimentos  
(Investment)

Saúde Financeira  
(Financial Health)

Volume de Transações  
(Transaction Volume)

Valor Médio de Transações  
(Transaction Value)

## Variáveis de Clusterização

**AvgMonthlyRevenueThousands:** quantidade média mensal do total do faturamento bruto mensal da agência (x1000)

**AvgMonthlyOperationalCostThousands:** quantidade média mensal do total de custo operacional mensal da agência (x1000)

**AvgMonthlyOperationalLossThousands:** quantidade média mensal do total de custo por perda operacional mensal da agência (x1000)

**AvgMonthlyEBITDA:** EBITDA médio mensal da agência

# Variáveis de perfil

Tamanho (Size)

Potencial de Negócios  
(Business Potential)

Crédito (Credit)

Investimentos  
(Investment)

Saúde Financeira  
(Financial Health)

**Volume de Transações  
(Transaction Volume)**

Valor Médio de Transações  
(Transaction Value)

## Variáveis de Clusterização

**AvgMonthlyATMPaymentsTransactions:** quantidade média mensal de transações do tipo pagamento realizados por caixa eletrônico na agência

**AvgMonthlyATMWithdrawTransactions:** quantidade média mensal de transações do tipo retirada/recebimentos realizados por caixa eletrônico na agência

**AvgMonthlyATMTransferTransactions:** quantidade média mensal de transações do tipo transferências realizados por caixa eletrônico na agência

**AvgMonthlyATMDepositTransactions:** quantidade média mensal de transações do tipo depósito bancário por caixa eletrônico realizados na agência

**AvgMonthlyATMTransactions:** quantidade média mensal de transações realizadas por caixa eletrônico da agência

**AvgMonthlyTellerPaymentsTransactions:** quantidade média mensal de transações do tipo pagamento realizados por caixa humano da agência

**AvgMonthlyTellerWithdrawTransactions:** quantidade média mensal de transações do tipo retirada realizados por caixa humano da agência

**AvgMonthlyTellerTransferTransactions:** quantidade média mensal de transações do tipo transferência realizadas por caixa humano da agência

**AvgMonthlyTellerDepositTransactions:** quantidade média mensal de transações do tipo depósito bancário realizados por caixa humano da agência

**AvgMonthlyTellerTransactions:** quantidade média mensal de transações totais realizadas por caixa humano da agência

**AvgMonthlyManagerTransactions:** quantidade média mensal do total de transações realizadas por gerente (pessoa física ou jurídica) da agência

# Variáveis de perfil

Tamanho (Size)

Potencial de Negócios  
(Business Potential)

Crédito (Credit)

Investimentos  
(Investment)

Saúde Financeira  
(Financial Health)

Volume de Transações  
(Transaction Volume)

Valor Médio de Transações  
(Transaction Value)

## Variáveis de Clusterização

**AvgMonthlySavingsAccountDeposit:** valor médio mensal de valores depositados por conta poupança pessoa física da agência

**AvgMonthlyPersonalLoanAmount:** valor médio mensal de valores de empréstimo mensal tomados por conta corrente pessoa física da agência

**AvgMonthlyBusinessLoanAmount:** valor médio mensal de valores de empréstimos mensal tomados por conta corrente pessoa jurídica da agência

**AvgMonthlyPersonalInvestmentsAmount:** valor médio mensal de volume aplicado em investimentos por conta corrente pessoa física da agência

**AvgMonthlyBusinessInvestmentsAmount:** valor médio mensal de volume aplicado em investimentos por conta corrente pessoa jurídica da agência

# Variáveis de perfil (1/3)

13

Feature Grouping	Features	Cluster 1 (Média)	Cluster 2 (Média)	Cluster 3 (Média)	Cluster 4 (Média)	Cluster 5 (Média)	Cluster 6 (Média)
Size	NumberTellerCapacity	↑ 8	↑ 8	→ 5	→ 5	↓ 2	↑ 8
Size	NumberManagerPersonalCapacity	↑ 3	↑ 3	→ 2	→ 2	↓ 1	↑ 3
Size	NumberManagerBusinessCapacity	↑ 3	↑ 3	→ 2	→ 2	↓ 1	↑ 3
Size	NumberATM	↑ 10	↑ 10	→ 6	→ 7	↓ 3	↑ 10
Size	NumberPersonalClientsTierA	↑ 1.586	↑ 1.209	→ 1.006	→ 977	↓ 443	→ 995
Size	NumberPersonalClientsTierB	↑ 3.195	↑ 3.185	→ 2.005	→ 2.013	↓ 900	→ 2.388
Size	NumberPersonalClientsTierC	↑ 3.983	↑ 4.036	↓ 2.495	→ 3.016	↓ 1.967	↑ 3.986
Size	NumberPersonalClientsTierD	↑ 5.554	↑ 5.538	↓ 3.493	→ 4.223	↓ 2.833	↑ 5.556
Size	NumberINSSClients	↑ 1.616	↑ 1.594	↓ 995	→ 1.201	↓ 800	↑ 1.599
Size	NumberSalaryAccounts	↑ 2.026	↑ 2.002	→ 1.251	→ 1.259	↓ 557	↑ 2.007
Size	NBusinessClientsTierA	↑ 120	→ 75	→ 75	→ 75	↓ 34	→ 75
Size	NBusinessClientsTierB	↑ 351	↑ 351	↓ 220	↓ 221	↓ 176	↑ 351
Size	NBusinessClientsTierC	↑ 717	↑ 719	↓ 452	→ 540	↓ 361	↑ 721
Size	NBusinessClientsTierD	↑ 1.365	↑ 1.377	↓ 848	→ 1.025	↓ 677	↑ 1.362
Size	BranchSizeSquareMeters	↑ 805	→ 601	↓ 502	→ 599	↓ 396	↑ 800

# Variáveis de perfil (2/3)

14

Feature Grouping	Features	Cluster 1 (Média)	Cluster 2 (Média)	Cluster 3 (Média)	Cluster 4 (Média)	Cluster 5 (Média)	Cluster 6 (Média)
BusinessPotential	AvgMonthlyIncomePersonalClientTierA	↑ 7.166	↑ 7.158	↓ 4.494	→ 5.401	↓ 3.579	↓ 3.610
BusinessPotential	AvgMonthlyIncomePersonalClientTierB	↑ 4.816	↑ 4.801	↓ 3.007	→ 3.633	↓ 2.390	↓ 2.403
BusinessPotential	AvgMonthlyIncomePersonalClientTierC	↑ 3.597	↑ 3.554	↓ 2.250	→ 2.688	↓ 1.808	↓ 1.780
BusinessPotential	AvgMonthlyIncomePersonalClientTierD	↑ 2.751	↑ 2.822	↓ 1.737	→ 2.089	↓ 1.430	↓ 1.408
BusinessPotential	AvgMonthlyINSSBenefitsAmount	↑ 3.125	↑ 3.134	↓ 1.953	→ 2.353	↓ 1.541	↓ 1.581
BusinessPotential	AvgMonthlySalaryValue	↑ 4.518	↑ 4.476	↓ 2.799	→ 3.373	↓ 2.239	↓ 2.250
BusinessPotential	AvgMonthlyEBITDABusinessClientTierA	↑ 798.971	↑ 795.564	↓ 499.852	→ 601.541	↓ 395.625	↓ 400.210
BusinessPotential	AvgMonthlyEBITDABusinessClientTierB	↑ 511.614	↑ 521.253	↓ 319.601	→ 383.358	↓ 247.501	↓ 257.656
BusinessPotential	AvgMonthlyEBITDABusinessClientTierC	↑ 234.016	↑ 235.871	↓ 150.661	→ 180.402	↓ 118.902	↓ 118.179
BusinessPotential	AvgMonthlyEBITDABusinessClientTierD	↑ 140.693	↑ 140.278	↓ 87.845	→ 105.593	↓ 70.371	↓ 70.210
Credit	AvgMonthlyManagerPersonalLoanTransactions	↑ 323	↑ 319	↓ 200	→ 241	↓ 162	→ 240
Credit	AvgMonthlyManagerBusinessLoanTransactions	↑ 480	↑ 482	↓ 301	→ 359	↓ 240	→ 363
Credit	AvgMonthlyPersonalCreditCardPaymentAmount	↑ 1.281	↑ 1.270	↓ 798	→ 962	↓ 646	↓ 796
Credit	AvgMonthlyBusinessCreditCardPaymentAmount	↑ 7.988	↑ 7.996	↓ 5.043	→ 6.014	↓ 4.003	↓ 5.050
Investment	AvgMonthlyManagerPersonalInvestmentTransactions	↑ 159	↓ 79	↓ 100	→ 120	↓ 79	→ 120
Investment	AvgMonthlyManagerBusinessInvestmentTransactions	↑ 238	↓ 120	↓ 150	→ 178	↓ 120	→ 181



# Variáveis de perfil (3/3)

15

Feature Grouping	Features	Cluster 1 (Média)	Cluster 2 (Média)	Cluster 3 (Média)	Cluster 4 (Média)	Cluster 5 (Média)	Cluster 6 (Média)
FinancialHealth	AvgMonthlyRevenueThousands	↑ 879.283	↑ 876.584	→ 662.031	→ 657.211	↓ 442.454	→ 662.513
FinancialHealth	AvgMonthlyOperationalCostThousands	↑ 239.104	↑ 241.601	↓ 67.424	→ 179.663	↓ 120.175	↑ 240.248
FinancialHealth	AvgMonthlyOperationalLossThousands	↑ 80.660	↑ 81.093	↓ 22.529	→ 60.066	↓ 39.405	↑ 80.205
FinancialHealth	AvgMonthlyEBITDA	↑ 556.841	↓ 280.023	→ 420.083	→ 418.740	↓ 282.672	↑ 558.893
TransactionValue	AvgMonthlySavingsAccountDeposit	↑ 875	↑ 881	↓ 549	→ 662	↓ 434	→ 662
TransactionValue	AvgMonthlyPersonalLoanAmount	↑ 1.785	↑ 1.824	↓ 1.124	→ 1.354	↓ 892	→ 1.336
TransactionValue	AvgMonthlyBusinessLoanAmount	↑ 39.720	↑ 39.972	↓ 24.962	→ 30.300	↓ 20.133	→ 30.075
TransactionValue	AvgMonthlyPersonalInvestmentsAmount	↑ 1.928	↓ 958	↓ 1.190	→ 1.429	↓ 958	↓ 972
TransactionValue	AvgMonthlyBusinessInvestmentsAmount	↑ 15.971	↓ 8.074	↓ 10.067	→ 12.029	↓ 7.991	↓ 10.022
TransactionVolume	AvgMonthlyATMPaymentsTransactions	↑ 39.790	↑ 40.054	→ 25.045	→ 29.838	↓ 20.211	↓ 11.320
TransactionVolume	AvgMonthlyATMWithdrawTransactions	↑ 32.125	↑ 32.004	→ 20.073	→ 24.120	↓ 16.178	↓ 9.202
TransactionVolume	AvgMonthlyATMTransferTransactions	↑ 23.988	↑ 23.959	→ 15.001	→ 18.054	↓ 12.030	↓ 6.742
TransactionVolume	AvgMonthlyATMDepositTransactions	↑ 30.270	↑ 30.528	→ 18.949	→ 22.674	↓ 15.039	↓ 8.744
TransactionVolume	AvgMonthlyATMTransactions	↑ 127.229	↑ 126.952	↓ 79.428	→ 95.167	↓ 62.548	↓ 63.846
TransactionVolume	AvgMonthlyTellerPaymentsTransactions	↑ 31.954	↑ 31.913	↓ 19.979	→ 24.026	↓ 15.684	↓ 15.913
TransactionVolume	AvgMonthlyTellerWithdrawTransactions	↑ 28.780	↑ 28.879	↓ 18.097	→ 21.618	↓ 14.400	↓ 14.483
TransactionVolume	AvgMonthlyTellerTransferTransactions	↑ 24.007	↑ 23.710	↓ 14.974	→ 18.048	↓ 11.927	↓ 12.099
TransactionVolume	AvgMonthlyTellerDepositTransactions	↑ 39.683	↑ 39.545	↓ 24.840	→ 30.008	↓ 19.976	↓ 20.011
TransactionVolume	AvgMonthlyTellerTransactions	↑ 126.287	↑ 123.783	↓ 77.713	→ 94.124	↓ 62.721	↓ 62.457
TransactionVolume	AvgMonthlyManagerTransactions	↑ 1.203	↑ 1.187	↓ 754	→ 899	↓ 596	→ 904



# Dispersão Média

Agrupamento de Variáveis	Cluster 1 (Dispersão)	Cluster 2 (Dispersão)	Cluster 3 (Dispersão)	Cluster 4 (Dispersão)	Cluster 5 (Dispersão)	Cluster 6 (Dispersão)	Avg Weighted Dispersion	Avg Dispersion
BusinessPotential	↓ 15%	↓ 14%	→ 23%	→ 19%	↑ 29%	↑ 29%	23%	22%
Credit	↓ 10%	↓ 10%	→ 16%	→ 13%	↑ 20%	→ 15%	15%	14%
FinancialHealth	↓ 10%	↓ 13%	↑ 24%	↓ 13%	↑ 20%	↓ 11%	19%	15%
Investment	↓ 10%	↑ 20%	→ 16%	↓ 13%	↑ 21%	↓ 13%	15%	15%
Size	↓ 11%	↓ 12%	→ 18%	↓ 17%	↑ 30%	↓ 13%	17%	17%
TransactionValue	↓ 10%	→ 14%	→ 16%	→ 13%	↑ 20%	→ 15%	15%	15%
TransactionVolume	↓ 10%	↓ 10%	→ 16%	↓ 13%	↑ 20%	↑ 25%	17%	16%
Qtde de agências	400	300	2500	700	500	800		
Valores Médios	↓ 11%	↓ 13%	→ 19%	↓ 15%	↑ 23%	→ 17%		

- Agrupamento de variáveis com menor dispersão: Credit e Transaction Value.
- Agrupamento de variáveis com maior dispersão: Financial Health e Business Potential.
- Agrupamento de variáveis com dispersão mediana: Size e Transaction Volume.
- Cluster com menor dispersão média: Cluster 1.
- Cluster com maior dispersão média: Cluster 5.

# Valores Normalizados (z-score)

	Cluster 1 (Média)	Cluster 2 (Média)	Cluster 3 (Média)	Cluster 4 (Média)	Cluster 5 (Média)	Cluster 6 (Média)	Valores Médios	Desvio Médio	Dispersão Média
Business Potential	↑ 171.127	↑ 171.891	↓ 107.420	→ 129.043	↓ 84.539	↓ 85.929	124.991	32.362	26%
Credit	↑ 2.518	↑ 2.517	↓ 1.586	→ 1.894	↓ 1.263	↓ 1.612	1.898	413	22%
Financial Health	↑ 438.972	↑ 369.825	↓ 293.017	→ 328.920	↓ 221.177	↑ 385.465	339.562	58.525	17%
Investment	↑ 199	↓ 100	↓ 125	→ 149	↓ 100	→ 150	137	29	21%
Size	↑ 1.423	↑ 1.381	→ 891	→ 1.011	↓ 610	↑ 1.324	1.107	269	24%
Transaction Value	↑ 12.056	↑ 10.342	↓ 7.578	→ 9.155	↓ 6.082	→ 8.613	8.971	1.547	17%
Transaction Volume	↑ 45.938	↑ 45.683	↓ 28.623	→ 34.416	↓ 22.846	↓ 20.520	33.004	9.008	27%

- Valores médios de agrupamento de variáveis por cluster
- Normalização dos valores médios através do z-score (utilizando média e desvio médio por agrupamento)

$$z = \frac{x_i - \bar{x}}{\sigma_x} \quad (4)$$

- Sendo  $\bar{x}$ , e  $\sigma_x$  respectivamente o valor médio, e o desvio padrão médio dos valores  $x_i$  por cluster.
- Valores normalizados pelo z-score agora podem ser comparados entre clusters, e entre agrupamentos de variáveis

# Valores Normalizados (z-score)

18

	Cluster 1 (Média)	Cluster 2 (Média)	Cluster 3 (Média)	Cluster 4 (Média)	Cluster 5 (Média)	Cluster 6 (Média)	Valores Médios	Desvio Médio	Dispersão Média
Business Potential	↑ 171.127	↑ 171.891	↓ 107.420	→ 129.043	↓ 84.539	↓ 85.929	124.991	32.362	26%
Credit	↑ 2.518	↑ 2.517	↓ 1.586	→ 1.894	↓ 1.263	↓ 1.612	1.898	413	22%
Financial Health	↑ 438.972	↑ 369.825	↓ 293.017	→ 328.920	↓ 221.177	↑ 385.465	339.562	58.525	17%
Investment	↑ 199	↓ 100	↓ 125	→ 149	↓ 100	→ 150	137	29	21%
Size	↑ 1.423	↑ 1.381	→ 891	→ 1.011	↓ 610	↑ 1.324	1.107	269	24%
Transaction Value	↑ 12.056	↑ 10.342	↓ 7.578	→ 9.155	↓ 6.082	→ 8.613	8.971	1.547	17%
Transaction Volume	↑ 45.938	↑ 45.683	↓ 28.623	→ 34.416	↓ 22.846	↓ 20.520	33.004	9.008	27%

	Cluster 1 (z-score)	Cluster 2 (z-score)	Cluster 3 (z-score)	Cluster 4 (z-score)	Cluster 5 (z-score)	Cluster 6 (z-score)
Business Potential	↑ 1,43	↑ 1,45	↓ -0,54	→ 0,13	↓ -1,25	↓ -1,21
Credit	↑ 1,50	↑ 1,50	↓ -0,76	→ -0,01	↓ -1,54	↓ -0,69
Financial Health	↑ 1,70	↑ 0,52	↓ -0,80	→ -0,18	↓ -2,02	↑ 0,78
Investment	↑ 2,13	↓ -1,29	↓ -0,42	→ 0,42	↓ -1,29	→ 0,45
Size	↑ 1,17	↑ 1,02	→ -0,80	→ -0,35	↓ -1,84	↑ 0,81
Transaction Value	↑ 1,99	↑ 0,89	↓ -0,90	→ 0,12	↓ -1,87	→ -0,23
Transaction Volume	↑ 1,44	↑ 1,41	↓ -0,49	→ 0,16	↓ -1,13	↓ -1,39

Qtde de agências	400	300	2500	700	500	800
------------------	-----	-----	------	-----	-----	-----

# Valores Normalizados (z-score)

	Cluster 1 (z-score)	Cluster 2 (z-score)	Cluster 3 (z-score)	Cluster 4 (z-score)	Cluster 5 (z-score)	Cluster 6 (z-score)
Business Potential	↑ 1,43	↑ 1,45	↓ -0,54	→ 0,13	↓ -1,25	↓ -1,21
Credit	↑ 1,50	↑ 1,50	↓ -0,76	→ -0,01	↓ -1,54	↓ -0,69
Financial Health	↑ 1,70	↑ 0,52	↓ -0,80	→ -0,18	↓ -2,02	↑ 0,78
Investment	↑ 2,13	↓ -1,29	↓ -0,42	→ 0,42	↓ -1,29	→ 0,45
Size	↑ 1,17	↑ 1,02	→ -0,80	→ -0,35	↓ -1,84	↑ 0,81
Transaction Value	↑ 1,99	↑ 0,89	↓ -0,90	→ 0,12	↓ -1,87	→ -0,23
Transaction Volume	↑ 1,44	↑ 1,41	↓ -0,49	→ 0,16	↓ -1,13	↓ -1,39
Qtde de agências	400	300	2500	700	500	800

Valores Médios	Desvio Médio	Dispersão Média
124.991	32.362	26%
1.898	413	22%
339.562	58.525	17%
137	29	21%
1.107	269	24%
8.971	1.547	17%
33.004	9.008	27%

- **Cluster 1** possui os **melhores** indicadores (características) de resultados (benchmark).
- Cluster 2 possui o segundo melhor conjunto de resultados.
- **Cluster 5** possui os **piores** indicadores (características) de resultados.
- Cluster 4 pode ser considerado o grupo homogêneo mediano.
- Cluster 3 é o segundo pior cluster com as piores características de resultados.
- Cluster 3 é o com maior população de agências.
- Relembrando: Cluster 5 com maior dispersão das variáveis de agrupamento.
- Relembrando: Cluster 1 com menor dispersão das variáveis de agrupamento.

# Grupos Homogêneos

20

Agrupamentos de Variáveis	Descrição do agrupamento	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Business Potential	Capacidade da agência em gerar negócios através de produtos bancários comuns, linhas de crédito, e investimentos.	●	●	●	●	●	●
Credit	Apetite por crédito via perfil dos clientes pessoa física, e pessoa jurídica da agência.	●	●	●	●	●	●
Financial Health	Saúde financeira da agência, baseado na análise contábil (média mensal) entre custo, receita, e perdas.	●	●	●	●	●	●
Investment	Apetite por investimento dos clientes pessoa física e pessoa jurídica da agência.	●	●	●	●	●	●
Size	Tamanho físico da agência (metragem), quantidade de clientes PF e PJ, níveis de transações, e quantidade de colaboradores.	●	●	●	●	●	●
Transaction Value	Valores médios por transação realizados na agência. É a média dos valores realizados por clientes PF e por PJ.	●	●	●	●	●	●
Transaction Volume	Quantidade média das transações realizadas na agência por seus clientes PF e PJ.	●	●	●	●	●	●

**Cluster 1 (400 agências):** *benchmark*, grandes em tamanho físico, e em quantidade de clientes pf e pj (empresas), alto potencial de negócios, alto apetite por crédito, boa saúde financeira, altos níveis transacionais tanto em volume quanto em valores médios, os clientes destas agências possuem também um perfil investidor.

**Cluster 2 (300 agências):** perfis similares ao *Cluster 1*, alto potencial de negócios, boa saúde financeira, alto apetite por crédito, agências grandes, e possuem altos níveis transacionais em volume e valores, porém os clientes das agências deste *cluster* não possuem perfil de investidores.

**Cluster 3 (2500 agências):** perfil de agências com baixa saúde financeira, pouco potencial de negócios, são agências de médio porte, possuem pouco apetite por crédito, os níveis transacionais são baixos tanto em volume quanto por valores, e os clientes destas agências não possuem perfil de investidores.

**Cluster 4 (700 agências):** perfil mediano quando comparado relativamente com os outros *clusters*, todas as suas características são medianas, ou seja são agências de médio porte, com níveis transacionais médios tanto em volume quanto em valores, possuem potencial de negócios, e saúde financeira medianas, os clientes destas agências possuem um perfil conservador em relação à investimentos.

**Cluster 5 (500 agências):** agências com os piores perfis, tendo os piores indicadores, com baixa saúde financeira, baixo potencial de negócios, são agências muito pequenas, com pouco apetite por crédito, baixos valores e volumes transacionais, e seus clientes não possuem perfil de investidores.

**Cluster 6 (800 agências):** grande porte em tamanho físico, e em quantidade de clientes, saúde financeira relativamente boa em relação aos piores clusters neste sentido, baixo potencial de negócios, pouco apetite por crédito, clientes não tem perfis de investidores, níveis medianos de transações em volumes, porém possuem baixos valores transacionados em média.

- Encontrados 6 grupos comparáveis reduzindo a dimensionalidade de 5.200 agências para 6 grupos comparáveis.
- Proposta de metodologia de análise de resultado de utilização do algoritmo K-Means (criação de aglomerados).
  - Utilização da dispersão média (desvio/média) como métrica de análise do resultado dos grupos homogêneos.
- Proposta de metodologia na criação de grupos homogêneos baseado no resultado do K-Means e em regras de negócio.
  - Utilização de variáveis de resultado de agências como variáveis de clusterização.
  - Agrupamento das variáveis de clusterização em grupos gerais de variáveis (auxílio na análise geral dos grupos).
  - Utilização do z-score (média, desvio) como métrica de análise geral entre os grupos.
- Proposta de análise dos valores “benchmark” na criação dos grupos homogêneos
  - Valores de dispersão média **17%** dentro dos grupos homogêneos.
- Grupos homogêneos apresentaram perfis destacados, com características de resultado características.



# Bibliografia

---

- MacQueen, J.B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1. University of California Press. pp. 281–297.
- Tang, G.; Tian, R.; Wu, B. (2022). An Overview of Clustering Methods in The Financial World. Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development Atlantis Press.
- Herrera-Restrepo, O.; Triantis, K.; Seaver, W.L.; Paradi, J.C.; Zhu, H.; Bank branch operational performance: A robust multivariate and clustering approach, Expert Systems with Applications, Volume 50, 2016, Pages 107-119, ISSN 0957-4174.
- Marques, B.P.; Alves, C.F. (2020) Using clustering ensemble to identify banking business models. Intell Sys Acc Fin Mgmt. 27: 66– 94.
- Domeniconi, C.; Gunopulos, D.; Ma, S.; Papadopoulos, D.; Yan, B. (2007) Locally adaptive metrics for clustering high dimensional data. Data Mining and Knowledge Discovery, Volume 14, Issue 1, pp 63–97.
- Sharahi, M.; Aligholi, M. (2015) Classify the Data of Bank Customers Using Data Mining and Clustering Techniques (Case Study: Sepah Bank Branches Tehran), J. Appl. Environ. Biol. Sci., 5(5) 458-464.
- T. E. Raghunathan, Synthetic data, Annual Review of Statistics and Its Application, 8, 129-140, 2021.
- Dankar, K.; Mahmoud, I. (2021) Fake it till you make it: guidelines for effective synthetic data generation, Applied Sciences 11.5: 2158.
- Hradec, J.; Craglia, M.; Di Leo, M.; De Nigris, S.; Ostlaender, N.; Nicholson, N. (2022) [Multipurpose synthetic population for policy applications](#), EUR 31116 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-53478-5.