

Relatório de Segmentação de Clientes – Portfólio SCR – Estado AL

Cientista de Dados – Projeto 3

Table of contents

1	1. Sumário Executivo	1
2	2. Introdução e Objetivo	2
3	3. Metodologia de Segmentação	2
3.1	3.1. Fonte de Dados	2
3.2	3.2. Justificativa do Algoritmo	3
3.3	3.3. Definição do Número de Clusters	3
4	4. Resultados: Perfis de Clientes	3
4.1	4.1 Cluster 1 – “Risco Emergente”	4
4.2	4.2 Cluster 2 – “Empresas Consolidadas e Estáveis”	4
4.3	4.3 Cluster 3 – “Alto Endividamento Estrutural”	4
5	5. Recomendações Estratégicas	5
6	6. Apêndice – Código R Completo	5

1 1. Sumário Executivo

A análise de clusterização aplicada ao portfólio de crédito PJ do SCR no estado de Alagoas identificou **3 perfis distintos de clientes**.

Os grupos mais relevantes são:

- **Cluster 1 – Risco Emergente:** empresas menores, maior propensão a inadimplência no curto prazo.
- **Cluster 2 – Empresas Consolidadas e Estáveis:** clientes sólidos, baixo risco e maior capacidade de expansão de crédito.
- **Cluster 3 – Alto Endividamento Estrutural:** empresas com forte concentração de operações longas e maior alavancagem.

Recomendações-chave:

- Cluster 1 → *monitoramento imediato e estratégias de renegociação*
- Cluster 2 → *cross-sell e expansão de crédito*
- Cluster 3 → *análise aprofundada de risco e revisão de garantias*

Este relatório detalha a metodologia, resultados e recomendações estratégicas derivadas da segmentação.

2 2. Introdução e Objetivo

O banco necessita segmentar seus clientes para melhorar:

- gerenciamento de risco,
- priorização de cobrança,
- concessão de crédito,
- oportunidades de venda,
- monitoramento preventivo de inadimplência.

O objetivo deste estudo é:

Identificar perfis naturais de clientes (clusters) no portfólio PJ utilizando técnicas de aprendizado não supervisionado.

3 3. Metodologia de Segmentação

3.1 3.1. Fonte de Dados

O dataset utilizado é o resultado do **Projeto 2**, contendo:

- 12.520 clientes PJ (observações)
- 31 variáveis, sendo:
 - numéricas padronizadas (z-score)
 - categóricas
 - variáveis derivadas de risco

Todos os dados foram previamente limpos, transformados e padronizados.

3.2 3.2. Justificativa do Algoritmo

O dataset é **misto**, composto por:

- Variáveis numéricas padronizadas
- Variáveis categóricas reais

Portanto, métodos tradicionais têm limitações:

Método	Aceita Numérico	Aceita Categórico	Pode ser usado aqui?
K-Means			
K-Modes			
K-Prototypes			Ideal

Assim, foi utilizado o algoritmo:

K-Prototypes (pacote `clustMixType`), que combina distância Euclidiana para variáveis numéricas e distância de Hamming para variáveis categóricas.

3.3 3.3. Definição do Número de Clusters

Foram avaliados:

- Método do Cotovelo (WSS)
- Índice de Silhueta Média
- Visualização PCA

Ambas as métricas indicaram que:

$k = 3$ foi o valor que melhor equilibrou coesão interna e separação entre grupos.

4 4. Resultados: Perfis de Clientes

Nesta seção apresentamos os perfis identificados (personas), com resumo estatístico por cluster.

4.1 4.1 Cluster 1 – “Risco Emergente”

Características predominantes:

- Menor capital social
- Maior incidência de vencimentoo recente
- Score de risco mais elevado
- Menor tempo de operação

Interpretação:

Empresas vulneráveis a choques e comportamentos de curto prazo.

Ação recomendada:

Monitoramento preventivo e revisão de limites.

4.2 4.2 Cluster 2 – “Empresas Consolidadas e Estáveis”

Características predominantes:

- Capital social elevado
- Porte médio/grande
- Baixa inadimplência
- Operações longas e estáveis

Ação recomendada:

Cross-sell e expansão de crédito.

4.3 4.3 Cluster 3 – “Alto Endividamento Estrutural”

Características predominantes:

- Operações longas com grande montante
- Alto grau de alavancagem
- Risco setorial relevante

Ação recomendada:

Acompanhamento de garantias e estudo de capacidade de pagamento.

5 5. Recomendações Estratégicas

Cluster	Persona	Estratégia Recomendada
1	Risco Emergente	Cobrança preventiva, limite reduzido
2	Consolidadas Estáveis	Ofertas de crédito e produtos adicionais
3	Alto Endividamento	Revisão de garantias, monitoramento próximo

6 6. Apêndice – Código R Completo

```
# PACOTES
library(dplyr)
library(ggplot2)
library(clustMixType)
library(factoextra)

# CARREGAR DADOS
df <- read.csv("dataset_microsegmentos_risco_AL_preprocessado.csv")

# IDENTIFICAR TIPOS
colunas_cat <- names(df)[sapply(df, is.character)]
colunas_num <- names(df)[sapply(df, is.numeric)]

# K-PROTOTYPES
set.seed(123)
modelo <- kproto(df[, c(colunas_num, colunas_cat)], k = 3)

df$cluster <- modelo$cluster

# MÉDIAS E MODAS POR CLUSTER
medias <- df %>% group_by(cluster) %>% summarise(across(all_of(colunas_num), mean))
modas <- df %>% group_by(cluster) %>% summarise(across(all_of(colunas_cat), ~names(sort(table(
# PCA VISUAL
pca <- prcomp(df[colunas_num])
df_pca <- data.frame(PC1 = pca$x[,1], PC2 = pca$x[,2], cluster = as.factor(df$cluster))
ggplot(df_pca, aes(PC1, PC2, color = cluster)) + geom_point(alpha=0.6)
```