

Aplicação do algoritmo K-Means na Classificação de Clientes de uma Distribuidora de Produtos de Petróleo

Aluno: Hamilton Batista Lima Sobrinho
Orientador: Professor Dr. Ricardo Rocha¹

¹ Departamento de Estatística, Universidade Federal da Bahia, Brasil

Resumo

A competitividade empresarial é um grande desafio enfrentado diariamente pelas empresas de todos os portes. Em meio a tanta disputa, faz-se necessário encontrar formas de manter e alavancar os negócios. A segmentação de clientes permite que a empresa se concentre em construir fortes relações com os seus consumidores mais valiosos, obtendo dessa forma melhores resultados. O objetivo desse trabalho é o agrupamento da base de clientes de uma distribuidora de derivados de petróleo, de modo a permitir ações comerciais diferenciadas para cada perfil de grupo encontrado. Para isto, será utilizado o algoritmo K-means, baseado no aprendizado de máquina não supervisionado.

Keywords:s Agrupamento de clientes; K-means; Aprendizado de Máquina Não Supervisionado.

1. Introdução

EM meio aos desafios enfrentados pelo acirramento da competitividade nos dias de hoje, a área comercial de uma empresa deve procurar constantemente formas de se reinventar, diferenciando-se dos concorrentes, de modo a atender melhor às necessidades dos seus clientes, e conseqüentemente atingir melhores resultados. Para permanecer nesse mercado cada vez mais exigente, é essencial que a organização reflita sobre a ferramenta da segmentação em sua área de vendas. As necessidades dos clientes mudam ao longo do tempo, e são cada vez mais específicas. Campanhas antes direcionadas para grandes grupos, hoje não trazem mais os mesmos bons resultados. Ao lidar com muitas pessoas, é difícil entender a necessidade de cada cliente individualmente, principalmente se esses clientes possuem maneiras diferentes de se relacionar com o vendedor. Por outro lado, é praticamente impossível entender em um nível individual cada um comprador, sendo muito difícil trabalhar com a realidade específica de cada um. A análise de agrupamento de dados é um dos principais problemas exploratórios tratados pela área de mineração de dados. Estas técnicas de agrupamento possibilitam a separação de grupos em grandes volumes de dados, tarefa nada fácil de ser realizada somente através da observação simples dos dados. Há décadas diversos algoritmos foram propostos para solucionar os mais variados problemas de agrupamento de dados. O algoritmo K-Means é um dos algoritmos mais conhecidos para esta tarefa. Tendo surgido na década de 1950, o K-Means pega alguns pontos da amostra de dados e coloca-os em grupos k, sendo que k é a quantidade de grupos que desejamos escolher. Os grupos k são definidos por um ponto central, como uma bandeira, e a partir daí, junta-se a esse ponto central um outro ponto, se ele estiver mais próximo a esse do que de algum outro ponto central adjacente.

2. Processos

As etapas deste trabalho são as seguintes:

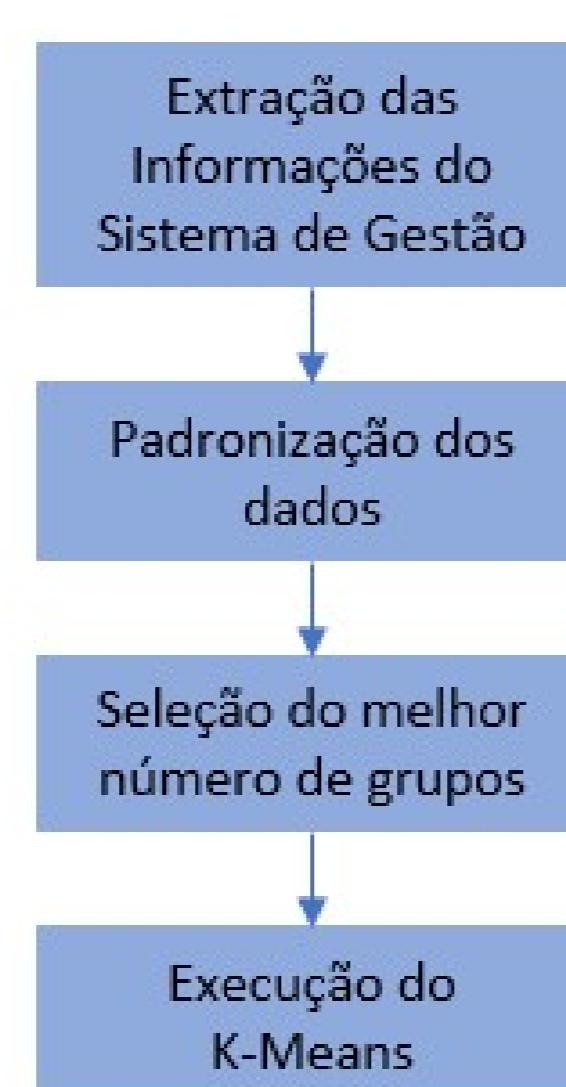


Figura 1: Etapas do Processo

Dados tratados após a padronização:

	Compras	Variância	Meses	Margem de Lucro	Frequencia	Pontualidade	Grupo Financeiro
1	-0.22876	-0.28102	0.01332	0.33645	-0.31156	-0.31282	0
2	-0.32558	-0.24434	0.01332	-0.41080	-0.48001	-0.23210	0
3	-0.39820	-0.13850	0.01332	0.18217	-0.60634	-0.11102	1
...
1471	-0.47081	0.14202	-1.20940	-0.12117	1.79394	0.79708	0
1472	-0.50885	-0.27626	-1.20940	-0.27662	-0.12508	-0.44736	0
1473	-0.49156	-0.22701	-1.20940	-1.34235	0.74720	-0.01012	1

Tabela 1: Dados após a Padronização

3. Metodologia

PARA a execução do algoritmo do K-Means, foram selecionadas as informações mais relevantes de vendas e dos clientes no banco de dados do Sistema de Gestão da companhia, a partir de comandos na linguagem SQL, sendo em seguida exportados no formato

Excel para a carga dos dados pelo Python. A fim de equalizar a ordem de grandeza das variáveis, as informações foram devidamente padronizadas. Para identificar a quantidade ideal de grupos para a aplicação do K-Means, foi utilizado o método Elbow, ou gráfico do cotovelo, através do qual, verificamos que a estabilização da variância acumulada (WSS) foi atingida a partir de 5 grupos. (Figura 2)

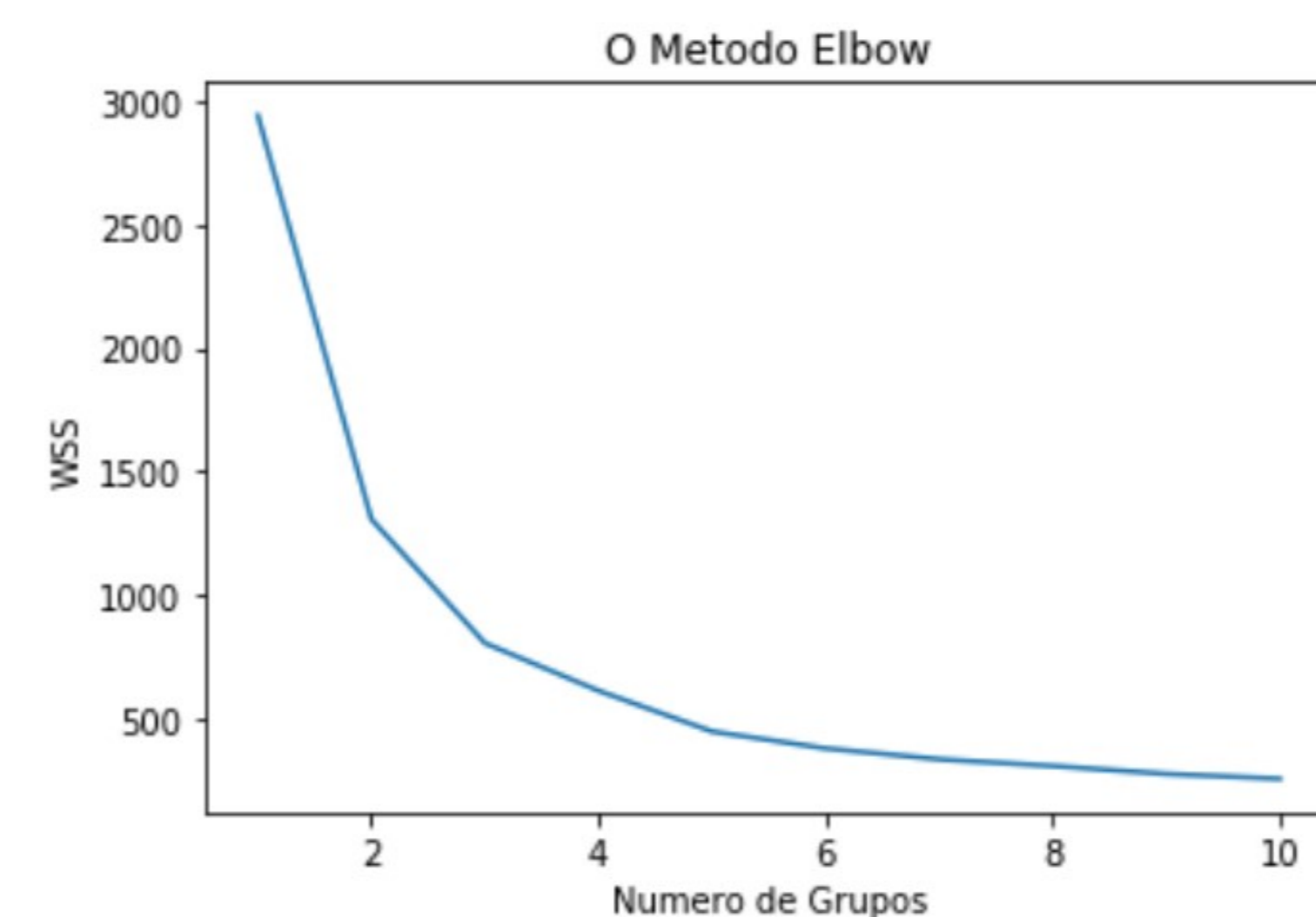


Figura 2: Etapas do Processo

4. Resultados

APÓS a execução do algoritmo com os dados tratados, o K-means revelou 5 diferentes grupos de clientes. Os grupos são apresentados na figura abaixo, identificados por cores diferentes.

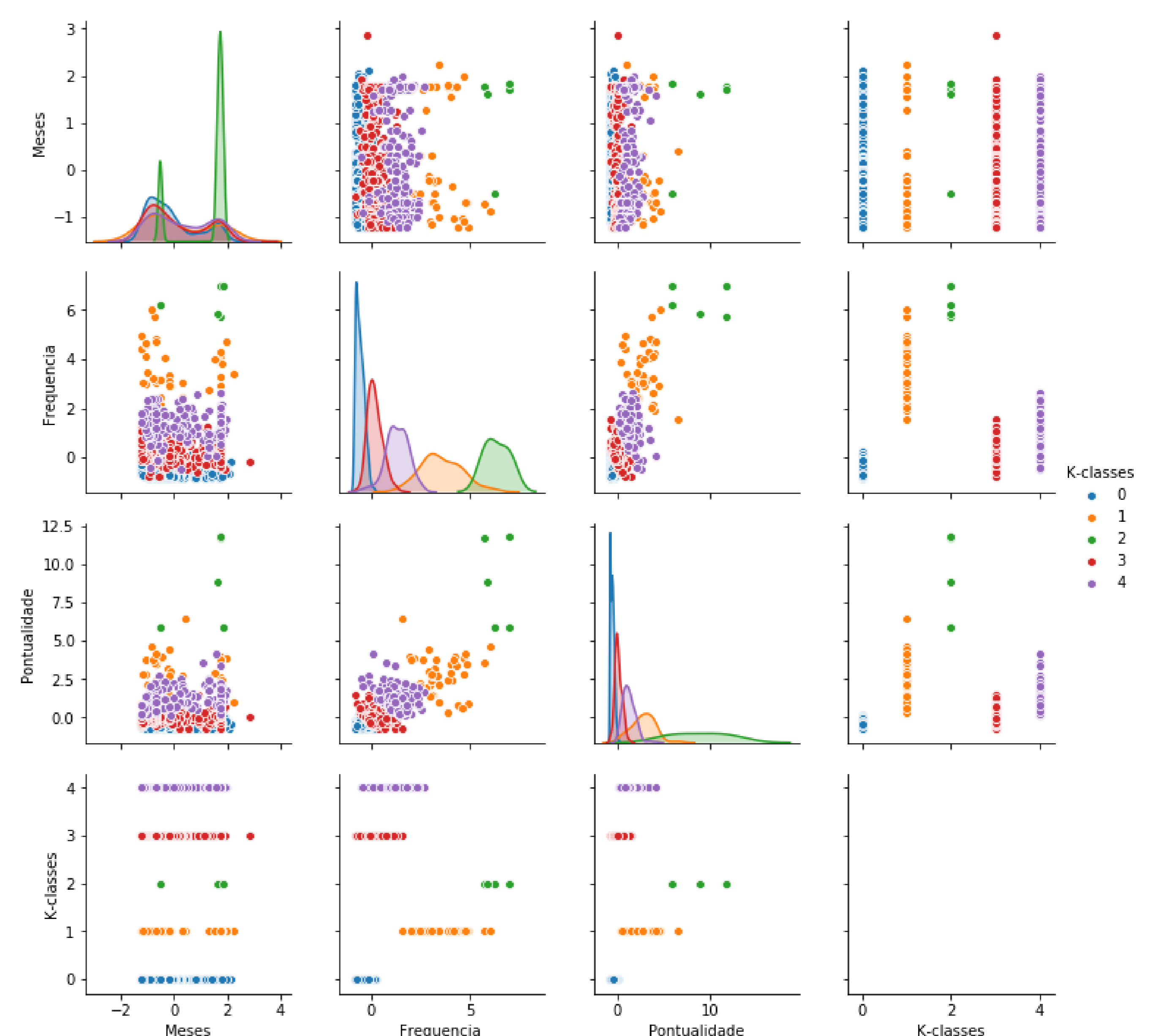


Figura 3: Resultados do K-Means

Conclusões

COM os resultados gerados pelo K-Means, pode-se descobrir agrupamentos antes não facilmente visualizados sem o auxílio da ferramenta. Estes agrupamentos, revelam afinidades entre os clientes que possibilitarão uma melhor exploração e direcionamento de ações pela área comercial da organização. O estudo destes nichos poderá melhorar o relacionamento com os seus clientes, gerar novas oportunidades de negócio e melhorar a competitividade da empresa neste cenário de alta concorrência.

Referências

- Tan, Pang (2009). *Introdução ao Data Mining*
- Foreman, John W. (2016). *Data Smart*
- <https://jakevdp.github.io/PythonDataScienceHandbook/05.11-k-means.html>
- <https://datatofish.com/k-means-clustering-python/>
- <https://www.kaggle.com/eriveltonguedes/7-clusteriza-o-k-means-erivelton/data>