

ESPECIFICAÇÕES DO PROJETO

Creating Customer Segments

Exploração dos dados

CRITÉRIO	ATENDEU ÀS ESPECIFICAÇÕES
Pergunta 1 Selecionando amostras	Três amostras diferentes dos dados são escolhidas, e o que elas representam é proposto com base na descrição estatística dos dados.
Pergunta 2 Relevância dos atributos	A pontuação do atributo removido foi corretamente calculada. A resposta justifica se o atributo removido é relevante.
Pergunta 3 Distribuição dos atributos	Atributos correlacionados são corretamente identificados e comparados com o atributo previsto. A distribuição dos dados para esses atributos é discutida.

Pré-processamento dos dados

CRITÉRIO	ATENDEU ÀS ESPECIFICAÇÕES
Pergunta 4 Detecção de valores aberrantes	Os valores aberrantes extremos são identificados, e discute-se se eles deveriam ser removidos. A decisão de remover quaisquer dados é corretamente justificada.
Dimensionamento de atributos	O código de dimensionamento de atributos, tanto para os dados como para as amostras, foi corretamente implementado.

Transformação de atributos

CRITÉRIO	ATENDEU ÀS ESPECIFICAÇÕES
Pergunta 5 PCA	A variância explicada total para duas e quatro dimensões dos dados do PCA é corretamente relatada. As primeiras quatro dimensões são interpretadas como uma representação dos gastos do cliente com justificativa.

Redução de dimensionalidade

O código do PCA foi corretamente implementado e aplicado, tanto para os dados dimensionados como para as amostras dimensionadas, no caso bidimensional.

Clustering

CRITÉRIO	ATENDEU ÀS ESPECIFICAÇÕES
Pergunta 6 Algoritmo de clustering	Os algoritmos GMM e k-means são comparados em detalhes. A escolha do aluno é justificada com base nas características do algoritmo e dos dados.
Pergunta 9 Estimativas de amostras	Amostras dos dados são corretamente relacionadas aos segmentos da clientela, e o grupo a que pertence cada ponto da amostra é discutido.
Pergunta 7 Criando grupos	Diversas pontuações são corretamente relatadas, e o número ótimo de grupos é escolhido com base na melhor. A visualização escolhida mostra o número ótimo de grupos baseado no algoritmo de clustering escolhido.
Pergunta 8 Recuperação dos dados	Os grupos representados por cada segmento da clientela são propostos com base na descrição estatística do conjunto de dados. Os códigos de transformação e dimensionamento inversos foi corretamente implementado e aplicado para o centro dos grupos.

Conclusão

CRITÉRIO	ATENDEU ÀS ESPECIFICAÇÕES
Pergunta 10 Teste A/B	O aluno identifica corretamente como um teste A/B pode ser feito com a clientela após uma mudança no serviço de distribuição.
Pergunta 11 Prevendo dados adicionais	O aluno discute e justifica como os dados de clustering podem ser usados em um modelo de aprendizagem supervisionada para fazer novas estimativas.
Pergunta 12 Comparando dados de clientes	Os segmentos da clientela e os dados em Channel são comparados. Os segmentos identificados pelos dados de Channel são discutidos, inclusive se essa representação é consistente com resultados anteriores.