

Estudo de Caso Avaliativo - Mineração de Dados & BI

Sumário

Introdução	1
Tarefas e Evidências Requeridas	1
Modelo de Estrutura do Relatório	4
Checklist de Evidências para Conferência do Grupo	6

Introdução

Você é parte do time de BI da rede de supermercados “SuperVida”, que deseja lançar um programa de alimentação saudável em parceria com ONGs locais. Seu desafio é usar técnicas de mineração de dados para:

- **Descobrir perfis de clientes com maior aderência** (ou potencial de aderência) ao programa,
- **Analisar padrões de compra e perfis demográficos**, e
- **Propor estratégias baseadas em evidências** para impulsionar o programa e melhorar a imagem da empresa.

Instruções Gerais

- Trabalho em duplas.
- **Vocês podem simular dados, completar dados reais (Kaggle) ou usar a estrutura de dados proposta.**
- **Entrega:** Relatório final (PDF), com códigos, gráficos e tabelas, além dos arquivos de dados e scripts usados.
- **Seja transparente:** Relate erros, dúvidas e decisões tomadas.
- **Evite dependência de IA para respostas prontas:** O foco está em raciocínio, exploração e justificativa.

Tarefas e Evidências Requeridas

1. Entendimento do Problema e dos Dados

Explique no relatório:

- O objetivo de negócio e perguntas centrais (“quem tem mais aderência ao programa?”).
- Limitações e potenciais vieses dos dados (por exemplo: dados ausentes, amostra não representativa, poucos atributos relevantes, enviesamento regional, etc).
- Identifique e justifique **quais colunas do banco de dados são mais relevantes** para a análise.
- **Evidência:** Texto explicativo (mínimo 1 página), com uma tabela descritiva dos atributos do dataset.

2. Pré-processamento de Dados

Detalhe, com código e print:

- **Como identificou e tratou valores ausentes, duplicados ou inconsistentes** (ex: 'grupo_caminhada', 'profissao', 'autoavaliacao_saude').
- **Criação de variáveis derivadas** (ex: faixas etárias, score de saúde, indicadores binários).
- **Conversão de dados categóricos para numéricos** (quando necessário).
- **Padronização de variáveis numéricas** (ex: z-score).
- **Como tratou outliers** (descreva o critério usado).
- **Evidência:** Mostrar trechos de código, prints da transformação dos dados e o novo arquivo CSV gerado.

3. Análise Descritiva (EDA)

No relatório e anexos:

- **Gráficos de barras e histogramas (com overlay/empilhamento):** comparar consumo de produtos naturais vs. ultraprocessados por faixa etária e escolaridade.
- **Tabela de contingência** mostrando relação entre profissão e consumo de produtos naturais.
- **Binning guiado:** Crie faixas etárias que melhor diferenciem os grupos.
- **Evidência:** Inclua gráficos, tabelas e uma breve interpretação dos resultados. Responda: "Existe algum grupo que se destaca?" "Há padrões inesperados?"

4. Agrupamento (Clustering) e Classificação

Comprovação com código e análise:

- **Aplique ao menos um algoritmo de clustering (ex: k-means):** justifique a escolha do número de clusters, mostre gráficos (scatter, silhouette, médias dos clusters).
- **Aplique ao menos dois algoritmos de classificação (ex: Árvore de Decisão, Naive Bayes, k-NN):** compare as acurácias usando matriz de confusão.
- **Destaque as variáveis mais relevantes para separar consumidores de produtos naturais/ultraprocessados.**
- **Evidência:** Prints dos outputs, gráficos dos clusters, comparação das acurácias/matriz de confusão, e um parágrafo interpretando os resultados.

5. Regras de Associação

Demonstre domínio do conceito:

- **Descubra regras do tipo "Se comprar produto X, também compra Y"** (ex: produtos naturais + orgânicos, snacks saudáveis etc.).

- **Interprete pelo menos duas regras relevantes usando suporte, confiança e lift.**
- **Avalie se algum perfil de cliente aparece em mais de uma regra forte.**
- **Evidência:** Códigos, tabela com as regras extraídas, explicação dos indicadores e interpretações práticas.

6. Tomada de Decisão de BI

Aponte estratégias com base nos achados:

- **Responda:** Quais perfis devem ser priorizados no programa?
- **Sugira ao menos uma ação concreta para cada grupo identificado** (ex: campanhas, eventos, promoções segmentadas).
- **Inclua ideias para envolver ONGs e grupos locais** (ex: caminhada, culinária).
- **Evidência:** Texto argumentativo, citando como os resultados dos modelos suportam as decisões. Tente incluir um pequeno fluxograma ou lista de “próximos passos”.

7. Rastreabilidade e Reprodutibilidade

Critério de excelência:

- Todos os códigos usados devem estar comentados e entregues (Jupyter Notebook, .py, ou equivalente).
- Todos os arquivos de dados usados ou gerados devem estar organizados na entrega.
- Inclua um "README" explicando como rodar o projeto.

Dicas:

- Relate dificuldades, escolhas metodológicas, e explique quando e por que um método funcionou melhor que outro (isso vale ponto!).
- Use exemplos e visualizações próprias - não apenas outputs de biblioteca!
- Busque trazer referências externas ao tema saúde/alimentação quando possível (ex: pesquisa de tendências, recomendações da OMS, hábitos regionais, etc).
- Caso use IA generativa, sempre valide manualmente e explique **o que você fez diferente ou adicional**.

Modelo de Estrutura do Relatório

Capa

- Título do trabalho
- Nomes dos integrantes da dupla
- Disciplina, professor, data

Sumário

1. Introdução

- Contexto do estudo de caso (objetivo, importância, relação com BI)
- Perguntas centrais a serem respondidas

2. Entendimento dos Dados

- Descrição do dataset (tabela de atributos, tipos de dados, tamanho da amostra)
- Limitações, possíveis vieses e desafios do conjunto de dados
- Justificativa dos atributos selecionados para análise

3. Pré-processamento de Dados

- Descrição dos problemas encontrados (valores ausentes, duplicados, inconsistências)
- Explicação das técnicas aplicadas (tratamento de valores, criação de variáveis, conversão de tipos, padronização, tratamento de outliers)
- Códigos utilizados (com comentários)
- Prints ou tabelas mostrando antes e depois do tratamento

4. Análise Descritiva (EDA)

- Gráficos (barras, histogramas, overlay)
- Tabelas de contingência
- Descrição dos principais padrões encontrados
- Interpretação dos gráficos e tabelas

5. Agrupamento e Classificação

- Descrição dos algoritmos utilizados (k-means, árvore de decisão, etc.)
- Justificativa da escolha dos parâmetros (ex: número de clusters)
- Gráficos dos clusters e médias dos grupos
- Métricas de avaliação (acurácia, matriz de confusão, relatório de classificação)
- Análise dos resultados: quais variáveis mais influenciaram, o que diferencia os grupos?

6. Regras de Associação

- Breve explicação do algoritmo Apriori ou similar
- Tabela com as regras encontradas (antecedente, consequente, suporte, confiança, lift)
- Interpretação das regras mais relevantes
- Comentários sobre aplicação prática dessas regras

7. Tomada de Decisão de BI

- Estratégias sugeridas para cada grupo/perfil identificado
- Ações concretas para impulsionar o programa de alimentação saudável
- Sugestões de envolvimento com ONGs, campanhas, etc.
- Justificativa das decisões, conectando com os resultados dos modelos
- Fluxograma ou tópicos de próximos passos

8. Conclusão

- Principais descobertas
- Dificuldades enfrentadas
- Limitações e sugestões para trabalhos futuros

9. Referências

- Fontes dos dados, artigos, sites, livros, etc.

10. Anexos

- Códigos completos e comentados (.py, .ipynb, etc.)
- Arquivos de dados utilizados/gerados
- Prints de outputs adicionais, gráficos, tabelas extras, etc.
- README de execução

Checklist de Evidências para Conferência do Grupo

Item	Feito?
1. Capa e identificação do grupo	
2. Introdução com perguntas centrais bem formuladas	
3. Descrição dos dados e tabela de atributos	
4. Discussão sobre limitações, vieses e desafios	
5. Explicação das etapas do pré-processamento (com prints/códigos)	
6. Geração de novo arquivo CSV após pré-processamento	
7. Gráficos de barras, histogramas e overlay	
8. Tabela de contingência e binning guiado	
9. Resultados e interpretações dos gráficos e tabelas	
10. Clusterização: justificativa do número de clusters, gráficos e médias	
11. Classificação: métricas de avaliação, matriz de confusão, comentários	
12. Tabela de regras de associação, interpretação de pelo menos duas regras	
13. Recomendações de BI claras e conectadas ao que foi descoberto	
14. Ações práticas sugeridas para a empresa e ONGs	
15. Todos os códigos entregues, bem comentados	
16. Todos os dados organizados (original, processado, outputs)	
17. README explicando como rodar tudo	
18. Conclusão reflexiva e sugestões futuras	
19. Referências bibliográficas e de dados	
20. Anexos organizados	